AD_____

Award Number: W81XWH-05-1-0314

TITLE: Early Life Processes, Endocrine Mediators and Number of Susceptible Cells in Relation to Breast Cancer Risk

PRINCIPAL INVESTIGATOR: Dimitrios Trichopoulos, M.D., Ph.D.

CONTRACTING ORGANIZATION: Harvard College
Boston, MA 02115

REPORT DATE: April 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-04-2008	Annual	15 MAR 2007 - 14 MAR 2008

4. TITLE AND SUBTITLE

Early Life Processes, Endocrine Mediators and Number of Susceptible Cells in Relation to Breast Cancer Risk

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0314

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dimitrios Trichopoulos, M.D., Ph.D.

E-Mail: dtrichop@hsph.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Harvard College
Boston, MA 02115

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Scope: To investigate the role of early life processes, endocrine mediators and number of susceptible cells on adult life breast cancer risk. Method: Five interlinked component projects covering the spectrum from endometrial to adult life. Progress report: Component projects 1 to 4 were officially launched July 2005. Component projects 5a and 5b were officially launched July 18, 2006. Tasks and subtasks to be performed were described in the submitted Statement of Work (SOW). Subtasks 1a, 1b, 2a, 2b, 3a, 3b, 3c, 4a, 4b, 4c, 5a, 5b have been completed. Subtasks 1c, 2c, 2d, 3d, 4d, 5c, 5d are ongoing. Subtasks 3e, 4e, 5f and 5g have been initiated. Subtasks 6a, 6b and 6c are being implemented. Major findings: (a) No substantial main effect of ESR1 and EGF on breast cancer risk, so that interaction with early life influences is unlikely (CP3) (b) Differences in the estradiol serum concentrations between the Boston and Shanghai pregnant women have been found to persist even after adjusting for proxies of plasma volume expansion (CP4). (c) Androgen levels have been found to be higher in cord blood samples from Chinese compared to Caucasian women, suggesting that elevated prenatal androgen exposure could mediate reductions in breast cancer risk (CP4). (d) Collagen I substrate has been found to be essential for the formation of mammospheres from epithelial cells found in cord blood (CP5). (e) Among term newborns with a normal-to-high birth weight, birth weight was found to be significantly positively associated with stem cell measurements, supporting a role of stem cell pool on cancer risk (CP5).

15. SUBJECT TERMS
Breast cancer, early life, mammary gland specific stem cells, hormones

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	47	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

INTRODUCTION

The aim of the project is to investigate the role of early life processes, endocrine mediators and number of susceptible cells on adult life breast cancer risk. Based on the hypothesis that breast cancer risk is a function of number of mammary gland cells at risk of transformation and that this number is largely modulated by perinatal events and conditions, five component projects have been initiated. The first three focus on perinatal characteristics, including immediate postnatal growth, in relation to mammary gland mass and breast cancer risk, whereas the last two explore the relation of pregnancy hormones with breast cancer risk and with cellular populations that are likely to have mammary stem cell potential. The five projects are interlinked and they address the hypothesis that growth and mammotropic hormones in perinatal life affect the number of susceptible mammary gland cells. This number is likely to be reflected in birth size and rate of postnatal growth that, in turn, represent intermediate steps and correlates of mammary gland mass and breast cancer risk in adult life. The progress on each component project (CP) will be reported separately to facilitate the reader.

BODY

CP1 "Association of growth during the first postnatal week with breast cancer risk in adult life"

CP1 PI: Prof. Anders Ekbom, Unit of Clinical Epidemiology, Dept. of Medicine, Karolinska Institutet/Karolinska University Hospital, SE-171 76 Stockholm, Sweden.

Timetable of research accomplishments of CP1 as outlined in the Statement of Work.

Task 1 *To investigate the association of growth during the first postnatal week with breast cancer risk in adult life:*

a. Retrieval of available birth records from 1,068 women with incident breast cancer and 2,727 control women. (Months 1-24)
b. Extraction of data on growth of newborns during the first postnatal week, as well as information on covariates to be used in the analysis. (Months 25-30)
c. Linkage of data on postanatal growth and perinatal covariates to cancer and mortality registries. (Months 31-36)
d. Data analyses. (Months 37-48)
e. Manuscript preparation and submission. (Months 49-60)

CP1 progress report

The retrieval of birth records (task 1a) is completed.

The extraction of data on growth of newborns (task 1b) is completed.

Linkage with cancer and mortality registries are ongoing and will be completed before July 1st 2008 according to schedule.

In accordance with the time plan, tasks 1d and e have not yet been initiated.

CP1 key research accomplishments

- The retrieval of birth records from 1,068 women with incident breast cancer and 2,727 control women is completed.
- The extraction of data on growth of newborns during the first postnatal week is completed.

CP1 reportable outcomes

No outcomes can be reported at this stage.

CP1 Conclusion

No major problems have been encountered and none are foreseen for the continuation of the implementation of the component project.

CP2: "Relation of perinatal characteristics and postnatal growth velocity with mammographic patterns in adult life"

CP2 PI: Prof. Per Hall, Dept. of Medical Epidemiology and Biostatistics, Karolinska Institutet, P.O. Box 281, SE-171 77 Stockholm, Sweden

Timetable of research accomplishments of CP2 as outlined in the Statement of Work.

Task 2 To investigate the relation of perinatal characteristics and postnatal growth velocity with mammographic patterns in adult life:

a. Retrieval of available birth records from 3,345 women with invasive breast cancer and 3,454 controls (these women are not the same with those to be studied in the context of task 1). (Months 1-48)
b. Retrieval of the available sequential mammographies of the women with breast cancer and the control women. (Months 1-48)
c. Linkage of mammographic data to perinatal characteristics and postnatal growth velocity. (Months 30-48)
d. Evaluation of mammographies through a computer-assisted grey-scale thresholding methods technique. (Months 25-36)
e. Data analyses. (Months 37-48)
f. Manuscript preparation and submission. (Months 49-60)

CP2 progress report

The retrieval of birth records (task 2a) ended in the fall of 2007. The retrieval of available mammograms (task 2b) is multi-step procedure. In a first step, images are identified and brought to the Department of Medical Epidemiology and Biostatistics at Karolinska. We have identified 45,000 images and brought them to the department. We are not able to digitize and measure all images and have therefore validated the concordance between left and right breast and the two major projections used in mammograms, cranio-caudal (CC) and medio-lateral-oblique (MLO), in a subset. There was a very high concordance for both right / left breast and MLO / CC mammographic projections. Nearly all examinations include MLO but not CC. Furthermore, most comparable projects use MLO and this projection has thus been chosen.

In a second step, all eligible images are digitized. This is an extremely time consuming endeavor and only 200 images could be processed in one day. We have had 2 persons working on digitizing and measuring images and have employed an additional person in order to use the full potential of the scanner. We have so far digitized images from 900 breast cancer patients and 600 controls out of a total of 2200 breast cancer patients and 2100 controls for which we have images.

In a third step, the density of each mammogram is measured. The evaluation of mammographies through a computer-assisted grey-scale thresholding methods technique (task 2d) will be done with the Cumulus software, a semi-automatic computer assisted technique where the evaluator indicates the area to be measured. The program is installed, and the responsible Swedish scientists have attended a course on the method given by the developers of the program, Norman Boyd and Martin Yaffe in Toronto in April 2007. A third person will take the course in April 2008.

We are collaborating with radiologists at Karolinska Hospital, as well as with Isabel dos Santos Silva and Julian Peto at London School of Hygiene and Tropical Medicine to assure good quality for the scanning and density reading methods. This group has long studied the relation of mammographic density with breast cancer risk and is highly experience in digitizing and evaluating density. We validated our measurements of 160 mammograms to those of the London group's for the same images and the agreement was high.

In accordance with the timetable of CP2, task 2c, e and f have not yet been initiated.

CP2 key research accomplishments

- All eligible and identified mammograms and birth characteristics have been retrieved.
- The image digitizing process is ongoing and running smoothly, although it has proven more time consuming than anticipated. We have so far digitized images from 900 breast cancer patients and 600 controls out of a total of 2200 breast cancer patients and 2100 controls for which we have images.
- Measurements of mammographic density from our group have been validated against those of an experienced British group, assuring an overall high quality of the measurements.

- Following the validation study, measurements of mammographic density have started.

CP2 reportable outcomes

At this stage, there are no reportable outcomes.

CP2 Conclusion

As outlined under "research accomplishments", the rate of retrieving information on birth characteristics and mammographic images is in line with what we anticipated and was finished in the fall of 2007. With the exception of a digitizing rate lower than initially anticipated, no major problems have been encountered and none are foreseen for the continuation of the implementation of this component project. In order to accelerate the process, an additional person has been hired, starting late March 2008.

CP3: "Interaction of perinatal characteristics with genes that are likely related to breast cancer risk"

CP3 PI: Prof. Per Hall, Dept. of Medical Epidemiology and Biostatistics, Karolinska Institutet, P.O. Box 281, SE-171 77 Stockholm, Sweden

Timetable of research accomplishments of CP3 as outlined in the Statement of Work.

Task 3 To investigate the possible interaction of perinatal characteristics with genes that are likely related to breast cancer risk:

a. Identification and selection of genes likely to be related to breast cancer risk, e.g. ESR1, AIB1, and the IGF family (Months 1-12)
b. Selection of "tagging" single nucleotiod polymorphisms (tSNPs). The choice of tSNBs aims at avoiding redundant genotyping. A good marker coverage is expected to be achieved by using approximately one SNP per 3 Kb. (Months 1-12).
c. Genotyping of the approximately 8 genes selected for the study (Months 13-36).
d. Data analyses. (Months 36-48)
e. Manuscript preparation and submission. (Months 49-60)

CP3 progress report

We have genotyped CHEK2, ESR1, EGF, HER2, ATM and 30 genes in the estrogen metabolizing pathway (including CYP191A and COMT). Since the aim is the study of interaction of genes with early life parameters with respect to breast cancer risk and mammographic density, data on ESR1 and EGF were first analyzed in relation to breast cancer risk, without any linkage to birth characteristics (Breast Cancer Res. 2008 Feb 14;10(1):R15). The study did not indicate a substantial main effect of ESR1 and EGF on breast cancer risk, so that interaction with early life influences is unlikely. Based on these results and in light of what has been published over the past 2-3 years in the international literature, we have chosen to also

focus on the estrogen metabolizing pathway. Results concerning genes and their possible interaction with early life exposures with respect to breast cancer risk and mammographic density will be studied in conjunction with the findings of CP2 as soon as the relevant data become available.

CP3 key research accomplishments

- Genotyping has been completed.
- Data on ESR1 and EGF were analyzed in relation to breast cancer risk, without any linkage to birth characteristics (Breast Cancer Res. 2008 Feb 14;10(1):R15)

CP3 reportable outcomes

- No substantial main effect of ESR1 and EGF on breast cancer risk, so that interaction with early life influences is unlikely.

CP3 Conclusion

As reported, tasks a, b and c have been completed according to schedule.

CP3 References

Einarsdóttir K, Darabi H, Li Y, Low YL, Li YQ, Bonnard C, Sjölander A, Czene K, Wedrén S, Liu ET, Hall P, Humphreys K, Liu J. ESR1 and EGF genetic variation in relation to breast cancer risk and survival. Breast Cancer Res. 2008 Feb 14;10(1):R15 [Epub ahead of print]

CP4: "Pregnancy hormones and perinatal breast cancer risk factors in Boston, USA and Shanghai, China"

CP4 co-PIs: Prof. Dimitrios Trichopoulos and Dr. Pagona Lagiou, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115,

Timetable of research accomplishments of CP4 as outlined in the Statement of Work.

Task 4 To study maternal and cord blood levels of components of the IGF system and adiponectin among Caucasian women in North America and Chinese women in Asia in conjunction to maternal anthropometry and birth size parameters:

a. Retrieval of the available stored cord blood and maternal serum samples from 304 pregnant Caucasian women in Boston, US and 335 pregnant Chinese women in Shanghai, China and transfer of these samples to the laboratory for hormone determination. (Months 1-6)

b. Conduct of laboratory assays for hormones (Months 7-24)
c. Linkage of maternal and newborn data to maternal and cord blood hormone levels (Months 24-30)
d. Data analyses (each of the measured hormones in maternal and cord blood to be studied in conjunction to maternal and newborn variables) (Months 31-48)
e. Manuscript preparation and submission. (Months 37-60)

CP4 progress report

With respect to baseline data, the database is already in place, as reported already in the 1st annual report. During the second reporting period, hormone determinations by the UMass Medical School, ILAT Steroid RIA Laboratory have been completed for 661 serum samples and 114 cord blood samples provided by pregnant women recruited in Boston, as well as for 762 serum samples and 131 cord blood samples provided by pregnant women recruited in Shanghai, China. Testosterone, IGF-1, IGFBP-3, and adiponectin have been measured on the maternal samples, while for cord blood samples, in addition to these hormones, estradiol, estriol, progesterone, SHBG and IGF-2 have been measured (measurements of these hormones in the maternal sera had already been conducted in the context of an earlier NIH study). During this third reporting period, data on hormone determinations in the serum and cord blood samples were merged with the baseline information on pregnant women and their newborn. The updated database contains information on: (i) baseline data and serum hormone measurements from 302 eligible women in Boston and 339 eligible women in Shanghai and (ii) baseline data and cord blood hormone measurements from 111 eligible newborn in Boston and 122 eligible newborn in Shanghai. Exploratory analyses have been conducted and two papers have been published. The first paper argues that measurement of the concentration of hormones and other biomarkers in pregnant subjects is influenced by plasma volume expansion (PVE) and suggests that PVE is a topic for consideration in population-based studies. The study compares the estradiol serum concentrations between the Boston and Shanghai pregnant women adjusting for proxies of PVE. Given that the difference persists, it appears that the difference is real, unless the used proxies are not sufficient markers of PVE (Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3). The second paper compares levels of cord blood hormones between Chinese and Caucasian and suggests that elevated prenatal androgen exposure among Chinese women could mediate reductions in breast cancer risk in the offspring (Cancer Epidemiol Biomarkers Prev. 2008;17:224-31).

Tasks 4a, 4b and 4c have been completed. Tasks 4d and 4e have been initiated.

CP4 key research accomplishments

- Merging of data on hormone determinations in the serum and cord blood samples with the baseline information on pregnant women and their newborn has been completed.
- Differences in the estradiol serum concentrations between the Boston and Shanghai pregnant women have been found to persist even after adjusting for proxies of plasma volume expansion (Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3).

- Androgen levels have been found to be higher in cord blood samples from Chinese compared to Caucasian women, suggesting that elevated prenatal androgen exposure could mediate reductions in breast cancer risk (Cancer Epidemiol Biomarkers Prev. 2008;17:224-31).

CP4 reportable outcomes

- Differences in the estradiol serum concentrations between the Boston and Shanghai pregnant women have been found to persist even after adjusting for proxies of plasma volume expansion (Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3).
- Androgen levels have been found to be higher in cord blood samples from Chinese compared to Caucasian women, suggesting that elevated prenatal androgen exposure could mediate reductions in breast cancer risk (Cancer Epidemiol Biomarkers Prev. 2008;17:224-31).

CP4 Conclusion

No major problems have been encountered and none are foreseen for the continuation of the implementation of the component project.

REFERENCES

Faupel-Badger JM, Hsieh CC, Troisi R, Lagiou P, Potischman N. Plasma volume expansion in pregnancy: implications for biomarkers in population studies. Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3.

Troisi R, Lagiou P, Trichopoulos D, Xu B, Chie L, Stanczyk FZ, Potischman N, Adami HO, Hoover RN, Hsieh CC. Cord serum estrogens, androgens, insulin-like growth factor-I, and insulin-like growth factor binding protein-3 in Chinese and U.S. Caucasian neonates. Cancer Epidemiol Biomarkers Prev. 2008;17:224-31.

CP5: "Breast stem cells and perinatal factors for breast cancer risk"

CP5 PI: Prof. Chung-Cheng Hsieh, University of Massachusetts Cancer Center, 55 Lake Avenue North, Worcester, MA 01655

Timetable of research accomplishments of CP5 as outlined in the Statement of Work.

Task 5 To investigate whether markers of mammary stem cells are associated with perinatal characteristics that are linked to breast cancer in later life:

a. Finalization of questionnaire for obtaining maternal and gestation characteristics. (Months 1-3)
b. Training of the study personnel on study procedures. (Months 1-6)
c. Subject recruitment and sample collection from a total of 250 pregnant women. (Months 7-42)
d. Conduct of laboratory assays for markers of stem cells (Months 7-45)
e. Conduct of laboratory assays for hormones (Months 10-48)
f. Data analyses. (Months 45-54)
g. Manuscript preparation and submission. (Months 49-60)

CP5 progress report

Clearance of component project 5 (a+b) was granted by the US Army Human Subjects Research Review Board (HSRRB) on July 17, 2006. Component project 5 was officially launched on July 18, 2006.

The underlying premises of this project are that 1) breast stem cells are the cell type that undergoes malignant transformation, 2) breast stem cells primarily arise during the fetal/perinatal period, and therefore the *in utero*/perinatal environment is a major determinant of the breast stem cell number in an individual, and 3) the greater the number of breast stem cells, the greater the likelihood of that one will undergo a genetic alteration that will be oncogenic. Previously we have shown that the concentration of hematopoietic stem cells in cord blood, serving as a surrogate for general stem cell potential, can be correlated to both perinatal levels of mitogens including estrogens and IGF-1, and to perinatal parameters such as birth weight, which have been linked to elevated breast cancer risk (Cancer Causes Control. 2004;15:517-30 and Cancer Res. 2005;65:358-363). Ideally, one would like to obtain some indicators of the levels of epithelial precursors cells in the perinatal environment, and determine if such cells might be an even better indicator of future breast cancer risk: this is the goal of Component Project 5, using umbilical cord blood as the perinatal cell source.

Since the launching of the project on July 18, 2006, the questionnaire for obtaining maternal and gestation characteristics has been finalized and study personnel have been trained on the study procedures (Task 5a and 5b).

Subject recruitment and sample collection started at the Tuft-New England Medical Center on November 16, 2006 (Task 5c). From November 16, 2006 to October 6, 2007, 132 women consented to participate in the study. Umbilical cord blood samples were collected from 87 eligible subjects. Of these samples, 83 samples were transferred to the stem cell laboratory within the requisite 24 hours of the childbirth. Of the transported samples, 80 were processed at the laboratory (three samples could not be processed: one due to low volume, the second due to inclement weather and the third due to arrival in late afternoon). It is noted that laboratory work was progressing according to plan until beginning of October 2007 when Dr. Todd Savarese

unexpectedly passed away. We are in the process of identifying and recruiting laboratory scientist(s) to continue the project.

In the past grant year, the Stem Cell Research Laboratory at the University of Massachusetts Medical School has made progress by processing 55 umbilical cord blood samples from normal pregnancies received from the Tuft-New England Medical Center (Task 5d). Of these, we have processed 53 samples using the protocol developed in the previous grant cycle. In this protocol, mononuclear cells (MNC) were extracted from the samples and analyzed by flow cytometry for hematopoietic cell markers (CD34+ and CD34+CD38-) and the epithelial adhesion molecule, EpCAM, a putative marker of breast epithelial stem cells. The EpCAM+ sub-populations obtained so far range from 0 to 0.08 cells/100 MNC, with a mean of 0.031 $\pm$ 0.025 (SD). We were also able to obtain *in vitro* data from 17 of these samples by culturing 2 X 10^5 purified EpCAM+ and EpCAM- MNC in a defined mammary epithelial cell growth medium (MEGM) in Collagen I-coated dishes and counting the number of "mammospheres" formed. While cultures of EpCAM- cells did not yield any cell clusters, the number of mammospheres obtained from EpCAM+ cells ranged from 0 to 474, showing a varying 'stem cell'-propagating potential among cord blood samples. Data analyses with flow cytometry are on-going and hormone assays for these samples will be undertaken when the number of samples is large enough to form efficient batches, thus minimizing cost and assuring optimal comparability of measurements. In addition, we are conducting experiments to explore the expression of other breast stem cell-associated markers on MNC extracted from umbilical cord blood. We have conducted flow cytometry analyses for cytokeratin, which is known to stain most epithelial-derived tissues, and for nestin, a marker recently reported to be expressed by mammary epithelia. While the staining with anti-cytokeratin (clone CAM5.2) confirmed a population of cells with epithelial characteristics in cord blood, the use of anti-nestin gave non-specific staining. We will continue to investigate whether other breast stem cell-associated markers, such as CD44 and CD24, are present in sub-population of cells in cord blood. We have also investigated the culturing of EpCAM+ MNC in MEGM in non-coated (no Collagen I) dishes but supplemented with bovine pituitary extract (BPE) and 10% fetal bovine serum (FBS). While the culturing of EpCAM+ MNC in un-supplemented MEGM, or supplemented with BPE resulted in senescence, the cultures in 10% FBS yielded only large single cells, but no cell clusters were observed. We thus conclude that a Collagen I substrate is essential to the formation of mammospheres from epithelial cells found in cord blood.

CP5 key research accomplishments

- Subject recruitment and sample collection, as well as laboratory assays for markers of stem cells are ongoing.
- Collagen I substrate has been found to be essential for the formation of mammospheres from epithelial cells found in cord blood.
- Analysis on previous data on hematopoietic stem cells in cord blood showed that, among term newborns with a normal-to-high birth weight, birth weight was found to be significantly positively associated with stem cell measurements, supporting a role of stem cell pool on cancer risk (Strohsnitter et al, 2008).

CP5 reportable outcomes

- Collagen I substrate has been found to be essential for the formation of mammospheres from epithelial cells found in cord blood.
- Analysis on previous data on hematopoietic stem cells in cord blood showed that, among term newborns with a normal-to-high birth weight, birth weight was found to be significantly positively associated with stem cell measurements, supporting a role of stem cell pool on cancer risk (Strohsnitter et al, 2008).

CP5 Conclusion

Component project 5 had been progressing according to plan until beginning of October 2007 when Dr. Todd Savarese unexpectedly passed away. We have provisionally identified a suitable successor and we expect the person to join the team on May 1, 2008.

CP5 References

Strohsnitter WC, Savarese TM, Low HP, Chelmow DP, Lagiou P, Lambe M, Edmiston K, Liu Q, Baik I, Noller KL, Adami H-O, Trichopoulos D, Hsieh CC. Correlation of umbilical cord blood hematopoietic stem and progenitor cell levels with birth weight: implications for a prenatal influence on cancer risk. Br J Cancer 2008;98:660-3.

Task 6: "Monitoring, coordination and fine-tuning of the five component projects"

Timetable of accomplishments of task 6 as outlined in the Statement of Work.

Task 6 To monitor, coordinate and fine-tune the five component projects:

a. Continuous monitoring and coordination of the research activities under the five component projects. (Months 1-60)
b. Compilation of annual overall project reports based on the component project-specific annual reports. (annually throughout the duration of the project) (Months1-60)
c. To coordinate the preparation and submission of manuscripts produced (Months 37-60)

Task 6 progress report

Monitoring and coordination of the five component projects has presented no problems. The key investigators have a long history of successful scientific collaboration, which continues in the context of the current project.

On May 9 2007, the project PI (DT) was awarded the Medal of Honor from the International Agency for Research on Cancer (IARC) of the WHO. His acceptance speech was dedicated to

research under the present Innovator award and led to a publication, co-authored by key contributors of the Innovator award (Int J Cancer. 2008;122:481-5).

KEY RESEARCH ACCOMPLISHMENTS

- The retrieval of birth records from 1,068 women with incident breast cancer and 2,727 control women is completed (CP1).
- The extraction of data on growth of newborns during the first postnatal week is completed (CP1).
- All eligible and identified mammograms and birth characteristics have been retrieved (CP2).
- The image digitizing process is ongoing and running smoothly, although it has proven more time consuming than anticipated. We have so far digitized images from 900 breast cancer patients and 600 controls out of a total of 2200 breast cancer patients and 2100 controls for which we have images (CP2).
- Measurements of mammographic density from our group have been validated against those of an experienced British group, assuring an overall high quality of the measurements (CP2).
- Following the validation study, measurements of mammographic density have started (CP2).
- Genotyping has been completed (CP3).
- Data on ESR1 and EGF were analyzed in relation to breast cancer risk, without any linkage to birth characteristics (Breast Cancer Res. 2008 Feb 14;10(1):R15) (CP3)
- Merging of data on hormone determinations in the serum and cord blood samples with the baseline information on pregnant women and their newborn has been completed (CP4).
- Differences in the estradiol serum concentrations between the Boston and Shanghai pregnant women have been found to persist even after adjusting for proxies of plasma volume expansion (Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3) (CP4).
- Androgen levels have been found to be higher in cord blood samples from Chinese compared to Caucasian women, suggesting that elevated prenatal androgen exposure could mediate reductions in breast cancer risk (Cancer Epidemiol Biomarkers Prev. 2008;17:224-31) (CP4).
- Subject recruitment and sample collection, as well as h laboratory assays for markers of stem cells are ongoing (CP5).
- Collagen I substrate has been found to be essential for the formation of mammospheres from epithelial cells found in cord blood (CP5).
- Analysis on previous data on hematopoietic stem cells in cord blood showed that, among term newborns with a normal-to-high birth weight, birth weight was found to be significantly positively associated with stem cell measurements, supporting a role of stem cell pool on cancer risk (Strohsnitter et al, 2008) (CP5).

REPORTABLE OUTCOMES

- No substantial main effect of ESR1 and EGF on breast cancer risk, so that interaction with early life influences is unlikely (CP3)
- Differences in the estradiol serum concentrations between the Boston and Shanghai pregnant women have been found to persist even after adjusting for proxies of plasma volume expansion (Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3) (CP4).
- Androgen levels have been found to be higher in cord blood samples from Chinese compared to Caucasian women, suggesting that elevated prenatal androgen exposure could mediate reductions in breast cancer risk (Cancer Epidemiol Biomarkers Prev. 2008;17:224-31) (CP4).
- Collagen I substrate has been found to be essential for the formation of mammospheres from epithelial cells found in cord blood (CP5).
- Analysis on previous data on hematopoietic stem cells in cord blood showed that, among term newborns with a normal-to-high birth weight, birth weight was found to be significantly positively associated with stem cell measurements, supporting a role of stem cell pool on cancer risk (Strohsnitter et al, 2008) (CP5).

CONCLUSIONS

Component project 5 had been progressing according to plan until beginning of October 2007 when Dr. Todd Savarese, a key lab scientist, unexpectedly passed away. We have provisionally identified a suitable successor and we expect the person to join the team on May 1, 2008.
With respect to component projects 1 to 4, no major problems have been encountered and none are foreseen for the continuation of the implementation of the project. At this stage, the early life origin of breast cancer appears well-supported and we have made progress in exploring the processes involved in the early life roots of breast cancer.

REFERENCES

Einarsdóttir K, Darabi H, Li Y, Low YL, Li YQ, Bonnard C, Sjölander A, Czene K, Wedrén S, Liu ET, Hall P, Humphreys K, Liu J. ESR1 and EGF genetic variation in relation to breast cancer risk and survival. Breast Cancer Res. 2008 Feb 14;10(1):R15 [Epub ahead of print]. (CP3)

Faupel-Badger JM, Hsieh CC, Troisi R, Lagiou P, Potischman N. Plasma volume expansion in pregnancy: implications for biomarkers in population studies. Cancer Epidemiol Biomarkers Prev. 2007;16:1720-3. (CP4)

Strohsnitter WC, Savarese TM, Low HP, Chelmow DP, Lagiou P, Lambe M, Edmiston K, Liu Q, Baik I, Noller KL, Adami H-O, Trichopoulos D, Hsieh CC. Correlation of umbilical cord blood hematopoietic stem and progenitor cell levels with birth weight: implications for a prenatal influence on cancer risk. Br J Cancer 2008;98:660-3. (CP5)

Trichopoulos D, Adami HO, Ekbom A, Hsieh CC, Lagiou P. Early life events and conditions and breast cancer risk: from epidemiology to etiology. Int J Cancer. 2008;122:481-5. (Task 6)

Troisi R, Lagiou P, Trichopoulos D, Xu B, Chie L, Stanczyk FZ, Potischman N, Adami HO, Hoover RN, Hsieh CC. Cord serum estrogens, androgens, insulin-like growth factor-I, and insulin-like growth factor binding protein-3 in Chinese and U.S. Caucasian neonates. Cancer Epidemiol Biomarkers Prev. 2008;17:224-31. (CP4)

APPENDIX

PDF copies of the indicated publications

Research article

Open Access

ESR1 and *EGF* genetic variation in relation to breast cancer risk and survival

Kristjana Einarsdóttir[1], Hatef Darabi[2], Yi Li[3], Yen Ling Low[3], Yu Qing Li[3], Carine Bonnard[3], Arvid Sjölander[2], Kamila Czene[2], Sara Wedrén[2], Edison T Liu[4], Per Hall[2], Keith Humphreys[2] and Jianjun Liu[3]

[1]Centre for Health Services Research, School of Population Health, University of Western Australia, 6009 Crawley, Perth, Western Australia
[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, 171 77 Stockholm, Sweden
[3]Human Genetics, Genome Institute of Singapore, Singapore 138672
[4]Cancer Biology, Genome Institute of Singapore, Singapore 138672

Corresponding author: Jianjun Liu, liuj3@gis.a-star.edu.sg

Received: 16 Oct 2007 Revisions requested: 22 Nov 2007 Revisions received: 25 Jan 2008 Accepted: 14 Feb 2008 Published: 14 Feb 2008

Breast Cancer Research 2008, **10**:R15 (doi:10.1186/bcr1861)
This article is online at: http://breast-cancer-research.com/content/10/1/R15

Abstract

Introduction Oestrogen exposure is a central factor in the development of breast cancer. Oestrogen receptor alpha (ESR1) is the main mediator of oestrogen effect in breast epithelia and has also been shown to be activated by epidermal growth factor (EGF). We sought to determine if common genetic variation in the *ESR1* and *EGF* genes affects breast cancer risk, tumour characteristics or breast cancer survival.

Methods We genotyped 157 single nucleotide polymorphisms (SNPs) in *ESR1* and 54 SNPs in *EGF* in 92 Swedish controls and selected haplotype tagging SNPs (tagSNPs) that could predict both single SNP and haplotype variation in the genes with an R^2 of at least 0.8. The tagSNPs were genotyped in 1,590 breast cancer cases and 1,518 controls, and their association with breast cancer risk, tumour characteristics and survival were assessed using unconditional logistic regression models, Cox proportional hazard models and haplotype analysis.

Results The single tagSNP analysis did not reveal association evidence for breast cancer risk, tumour characteristics, or survival. A multi-locus analysis of five adjacent tagSNPs suggested a region in *ESR1* (between rs3003925 and rs2144025) for association with breast cancer risk ($p = 0.001$), but the result did not withstand adjustment for multiple comparisons ($p = 0.086$). A similar region was also implicated by haplotype analyses, but its significance needs to be verified by follow-up analysis.

Conclusion Our results do not support a strong association between common variants in the *ESR1* and *EGF* genes and breast cancer risk, tumour characteristics or survival.

Introduction

Breast cancer is the most common cancer in women overall worldwide. Oestrogen exposure is a central factor in the development and progression of this cancer [1-3] and its effects on the breast epithelium is primarily mediated by oestrogen receptor alpha (ESR1) [4]. In addition to being activated by oestrogen, the ESR1 protein can be activated by growth factors such as epidermal growth factor (EGF) [3], which acts as a potent mitogen for epithelial cells, including mammary epithelia [5]. Variation in the *ESR1* (MIM 133430) and *EGF* (MIM 131530) genes affecting the function or expression of their respective proteins could thus potentially affect the risk of developing breast cancer, characteristics of the tumour or the risk of dying from the disease.

With regard to breast cancer risk or survival, a number of single nucleotide polymorphisms (SNPs) have been studied in the *ESR1* gene, yet none have previously been investigated in the *EGF* gene. As far as we are aware, no attempt to capture the common genetic variation in the *ESR1* gene in its entirety has yet been published. One group, who genotyped 17 SNPs in the *ESR1* gene, found a decreased risk of breast cancer for carriers of three common haplotypes in the gene and an

EGF = epidermal growth factor; ESR1 = (o)estrogen receptor 1; HWE = Hardy-Weinberg equilibrium; LD = linkage disequilibrium; MAF = minor allele frequency; NPI = Nottingham Prognostic Index; PLEM = partition ligation expectation maximisation; SNP = single nucleotide polymorphism; tagSNP = tagging single nucleotide polymorphism; TNM = tumour, nodes, metastasis.

increased risk for carriers of one common haplotype [6]. We genotyped 157 SNPs in *ESR1* and 54 SNPs in *EGF* using a population-based case-control study, which included 1,590 breast cancer cases and 1,518 controls. We selected haplotype-tagging SNPs (tagSNPs) spanning the *ESR1* and *EGF* genomic regions and assessed their association with breast cancer risk, the Nottingham Prognostic Index (NPI) and breast cancer survival.

Patients and methods

Parent breast cancer study

The study base included all Swedish-born women between 50 and 74 years of age and resident in Sweden between October 1993 and March 1995. During this period, all breast cancer cases were identified at diagnosis through the six regional cancer registries in Sweden. Controls were randomly selected from the Swedish Registry of Total Population to match the cases in 5-year age strata. Of the eligible cases and controls, 3,345 (84%) breast cancer cases and 3,454 (82%) controls participated in this initial questionnaire-based study.

Present breast cancer study

From the parent study, we randomly selected 1,500 breast cancer cases and 1,500 age- and frequency-matched controls among the postmenopausal participants without any previous malignancy (except carcinoma *in situ* of the cervix or non-melanoma skin cancer). With the intention of increasing statistical power in subgroup analyses, we further selected all remaining breast cancer cases and controls that had used menopausal hormones (oestrogen alone or any combination of oestrogen and progestin) for at least 4 years. We also included all remaining participants with self-reported diabetes mellitus. In total, we selected 1,801 breast cancer cases and 2,057 controls.

Following informed consent, participants donated whole blood. For deceased cases and those cases that declined to donate blood but consented to our use of tissue, we collected archived paraffin-embedded, non-cancerous tissue samples. We acquired 70% of the requested tissue samples; the main reason for non-participation was unwillingness or lack of time at the respective pathology department to provide the tissue blocks. In total, we obtained blood samples and archived tissue samples for 1,321 and 275 breast cancer patients, respectively, and blood samples for 1,524 controls. Population-based participation rates (taking into account the proportion that did not participate in the parent questionnaire study) were 75% and 61% for the cases and controls, respectively.

We extracted DNA from 4 ml of whole blood using the QIAamp DNA Blood Maxi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. From non-malignant paraffin-embedded tissues, DNA was extracted using a standard phenol/chloroform/isoamyl alcohol protocol [7]. We successfully isolated DNA from 1,318 (blood) and 272 (tissue) breast cancer patients and 1,518 controls. We randomly selected 92 out of the 1,518 controls to be used for linkage disequilibrium characterisation and haplotype reconstruction of the *ESR1* and *EGF* genes.

This study was approved by the Institutional Review Boards in Sweden and at the National University of Singapore.

SNP markers and genotyping

We selected SNPs in the *ESR1* and *EGF* genes and their 20 kb flanking sequences from dbSNP (build 124, [8]) and Celera databases, aiming for an initial marker density of at least one SNP per 5 kb. The *Pvu*II (rs2234693), *Xba*I (rs9340799), codon 243 (rs4986934) and codon 325 (rs1801132) variants were selected from the literature and added to our SNP selection. SNPs were genotyped using the Sequenom primer extension-based assay (San Diego, CA, USA) and the BeadArray system from Illumina (San Diego, CA, USA) following the manufacturers' instructions. All genotyping plates included positive and negative controls, DNA samples were randomly assigned to the plates, and all genotyping results were generated and checked by laboratory staff unaware of case-control status. Only SNPs where more than 85% of the samples gave a genotype call were analysed further. As quality control, we genotyped 200 randomly selected SNPs in the 92 control samples using both the Sequenom system and the BeadArray system. The genotype concordance was $>99.5\%$, suggesting high genotyping accuracy.

Linkage disequilibrium characterisation and tagSNP selection

We genotyped a dense set of SNPs in the *ESR1* and *EGF* genes in the 92 controls (Supplementary Tables 1 and 2 in Additional File 1, respectively). We identified regions of linkage disequilibrium (LD) and selected tagSNPs. We produced LD plots of the D' and R^2 values for *ESR1* and *EGF* (Supplementary Figures 1 and 2 in Additional File 1, respectively) using the *LDheatmap* function in the statistical software R [9]. We reconstructed haplotypes using the partition ligation expectation maximisation (PLEM) algorithm [10] implemented in the *tagSNPs* program [11] and selected tagSNPs based on the R^2 coefficient, described previously (equation (1) in [12]. In our case this is the squared correlation between the true number of haplotypes (defined across all SNPs typed in the 92 controls) and the number of copies of haplotypes predicted as being carried, based on the tagSNPs. The R^2 coefficient in [12] can also be used for measuring association between the genotypes of all SNPs typed in the 92 controls and the genotypes predicted on the basis of knowing the tagSNPs only. We chose tagSNPs so that common SNP genotypes (minor allele frequency ≥ 0.03) and common haplotypes (frequency ≥ 0.03) were predicted with $R^2 \geq 0.8$ [13]. The well studied *Pvu*II (rs2234693), *Xba*I (rs9340799), codon 243 (rs4986934) and codon 325 (rs1801132) variants were included as tagSNPs. In order to evaluate our tagSNPs' performance in

capturing unobserved SNPs within the genes and to assess whether we needed a denser set of markers, we performed a SNP-dropping analysis [12,14]. In brief, each of the genotyped SNPs was dropped in turn and tagSNPs were selected from the remaining SNPs so that their haplotypes predicted the remaining SNPs with an R^2 value of 0.85. We then estimated how well the tagSNP haplotypes of the remaining SNPs predicted the dropped SNP, an evaluation that can provide an unbiased and accurate estimate of tagSNP performance [12,14].

There were 19 SNPs upstream of the first tagSNP (TAG1) in *ESR1* (Supplementary Table 1 in Additional File 1). Of the 19, 12 were either not polymorphic or had a minor allele frequency (MAF) of less than 3%. The remaining seven SNPs in this area were included in our LD identification and tagSNP selection analysis. Hence, all polymorphic SNPs with a MAF ≥ 3% far 5' upstream of *ESR1* were captured by our tagSNPs.

Breast tumour characteristics and follow-up

We retrieved information on date and cause of death until 31 December 2003 from the Swedish Causes of Death Registry and on date of emigration from the Swedish National Population Registry. Follow-up time began at date of diagnosis and ended on 31 December 2003, or at date of death or emigration, whichever came first.

We collected information on tumour size, lymph node involvement, and grade (tumour differentiation) from medical records and calculated the Nottingham Prognostic Index (NPI) using the following formula:

NPI = 0.2 × size [in cm] + 1 × nodal stage [1, 2, or 3] + 1 × grade [1, 2, or 3] [15]

Nodal stage was defined as 1 if there were no lymph node metastases, 2 for a total of 1–3 metastatic nodes, and 3 for more than 3 metastatic nodes. A tumour of high differentiation was assigned grade 1, a tumour of intermediate differentiation grade 2, and a low differentiated tumour was assigned grade 3. We categorised the NPI into two groups: ≤ 4 or > 4. Four is the mean NPI value of the present study. It has also been shown that breast cancer survival decreases rapidly for NPI above 4 [15].

Statistical analyses

We applied unconditional logistic regression models for assessing the association between *ESR1* and *EGF* tagSNPs and risk of breast cancer (case-control analysis) or the NPI (case only analysis). Adjusting for age (in 5-year age groups) did not affect our results. We estimated the hazard ratio of death due to breast cancer in relation to the genes' tagSNP using Cox proportional hazards models. The tagSNPs were included as covariates in the models either one at a time or in groups of five (codominant main effects only). The latter method was used for detection of association with haplotypes. Although it does not require resolution of gametic phase, tests based on such models can be powerful within regions of strong LD [16]. Likelihood ratio tests were used to generate p values for comparing models with or without covariates. We made adjustments to our test results to account for multiplicity. We did so for each outcome (risk, NPI, and survival) separately. We used a permutation-based approach that controls the family-wise error rate (probability of rejecting one or more true null hypotheses of no association). This is based on the permutation step-down procedure of Westfall and Young [17] and takes into account the dependence structure of the polymorphisms/hypotheses. We also assessed association between groups of haplotypes and breast cancer risk using three approaches (each of which resolve gametic phase). We used the logistic regression expected haplotype dosage approach of [11], combining rare haplotypes. Since there is no biological reason to cluster haplotypes on the basis of their frequency, we also employed a Bayesian association mapping approach [18] that clusters haplotypes according to their allelic similarity, and a sliding-window approach described by Li *et al*. [19].

To estimate power in the risk component of the study, we used a method described by Chapman *et al*. [20], which assumes co-dominant effects at an unobserved locus. To calculate power for log-additive effects in the survival component of the study, we used the Quanto program [21] in a similar manner as Manolio *et al*. [22]. Analyses were performed using the statistical software R or the SAS system (v. 9.1, SAS Institute Inc., Cary, NC, USA). Because lifestyle and reproductive breast cancer risk factors are unlikely to cause genetic variation in the genes, we thus did not adjust for them in the analyses.

Results

Characteristics of participants

Table 1 shows selected characteristics of the cases and controls included in the parent questionnaire-based study and the current genetic study. Long-term users of menopausal hormone therapy and women with self-reported diabetes mellitus were oversampled in the current study. Most other characteristics were statistically significantly different between cases and controls and reflected established associations.

More case-related information has been provided in our previous work [23]. The breast cancer cases that participated in our study via tissue sample donation were on average 1.5 years older ($p = 0.0003$) than the cases that donated blood. The former group was also more likely to have been diagnosed with TNM (tumour, nodes, metastasis) stage 2 or more advanced cancers ($p < 0.0001$). Since no significant differences in genotype frequencies within TNM stage 1, TNM stage 2 and TNM stages 3 and 4 were evident between the

Table 1

Selected characteristics of the cases and controls participating in the present and parent breast cancer study

	Present		Parent	
Characteristic	No. of cases/controls	Cases/controls	No. of cases/controls	Cases/controls
		Mean:		Mean:
Age (years)	1,590/1,518	63.4/63.1	2,817/3,111	63.4/64.3
Age at menopause (years)	1,580/1,505	50.4/50.0	2,802/3,093	50.4/50.0
Recent BMI (kg/m^2)[a]	1,581/1,497	25.8/25.5	2,802/3,065	25.8/25.5
Age at first birth (years)	1,352/1,370	25.4/24.7	2,373/2,753	25.3/24.6
Parity	1,590/1,518	1.8/2.2	2,817/3,110	1.8/2.1
Duration of menopausal hormone use (years)		Percentage:		Percentage:
0	1,058/1,086	67.2/72.7	1,978/2,467	71.4/80.8
< 4	206/190	13.1/12.7	405/330	14.6/10.8
≥ 4	311[b]/217[b]	19.8[b]/14.5[b]	389/256	14.0/8.4
Self-reported diabetes mellitus (yes/no)	1,588/1,402	9.0[b]/7.8[b]	2,810/2,652	6/6.1
Family history (yes/no)[c]	1,551/1,380	16.1/9.3	2,744/2,607	16.0/9.2
High NPI (≤ 4/> 4)	975/-	55.7/-		-/-

BMI, body mass index; NPI, Nottingham Prognostic Index. [a] At 1 year prior to diagnosis. [b] Long-term users of menopausal hormones and women with diabetes mellitus were oversampled. [c] Family history is defined as having at least one first degree relative with breast cancer.

two groups of cases, this difference is unlikely to be a cause for concern.

Genotyping, LD pattern and coverage

The genotyping results and SNP coverage in the *ESR1* and *EGF* genes are summarised in Table 2. A dense set of SNPs in the *ESR1* and *EGF* genes were genotyped in 92 randomly selected controls (Supplementary Table 1 (*ESR1*) and Supplementary Table 2 (*EGF*) in Additional File 1), and only the SNPs that were in Hardy-Weinberg equilibrium ($p > 0.01$) and that were at least 3% in minor allele frequency among the 92 controls were included in LD analysis and tagSNP selection (Table 2). LD plots created from the SNPs included in our study are shown in Supplementary Figures 1 (*ESR1*) and 2 (*EGF*) in Additional File 1. Using the SNP dropping method [14], we found that the tagSNPs selected from the included SNPs could efficiently capture non-genotyped SNPs in the genes (Table 2).

Association analyses

We selected 52 tagSNPs in *ESR1* and 15 tagSNPs in *EGF* that could predict the included SNPs and their haplotypes with an R^2 of at least 0.8. The tagSNPs were genotyped in all cases and controls (Supplementary Table 3 in Additional File 1), but seven tagSNPs in *ESR1* and one tagSNP in *EGF* could not be genotyped in the cases that participated via tissue sample donation.

ESR1

For each outcome (breast cancer risk, NPI and breast cancer survival), we first tested the association of each tagSNP and then performed a haplotype analysis using a logistic regression sliding-window approach where five adjacent tagSNPs were analysed together (without resolution of gametic phase). The results are summarised in Figure 1 and Supplementary Table 4 in Additional File 1. Analysis of the 52 tagSNPs in *ESR1* (including the *Pvu*II, *Xba*I, codon 243 and codon 325 SNPs) did not reveal any association with breast cancer risk, NPI or breast cancer survival whose statistical significance withstood multiple testing correction. The strongest signal of association with breast cancer risk was obtained by the window analysis including TAGs 26–30 ($p = 0.001$ and $p = 0.086$, before and after correction for multiple testing). Within the region, there were seven common haplotypes that accounted for 92% of the chromosomes. Including the expected dosages of common haplotypes and the rare haplotypes (combined into a single variable) as covariates in a logistic regression model with the most common haplotype as reference, gave a global p (likelihood ratio test) of 0.0493 in relation to breast cancer risk (Table 3).

We also explored a sliding-window analysis of haplotypes using a variable window size. Three haplotypes within the region from tagSNP 18 to tagSNP 27 were implicated, showing frequency differences between cases and controls (Table 4). The significance of the frequency differences was, however, not clear, given the large number of haplotypes being searched in both fixed- and variable-sized sliding-window anal-

Table 2

Summary statistics on genotyping results and SNP coverage in *ESR1* and *EGF* for 92 Swedish controls

Summary statistics	*ESR1*	*EGF*
Number of successfully genotyped SNPs	228[a]	104[b]
Number of polymorphic SNPs	184	66
Number of common SNPs[c]	165	55
Number of SNPs deviating from HWE[d]	8	1
Number of SNPs included in study	157	54
Gene size (kb)	295.7	99.4
Sequence coverage of included SNPs (kb)	335.1	145.5
Mean spacing between included SNPs (kb)	2.1	2.7
Median spacing between included SNPs (kb)	1.8	2.3
Number of tagSNPs selected	52	15
Average tagSNP prediction of common SNPs included in study (R^2)[c]	0.998	0.987
Coverage evaluation[e]		
Average prediction of dropped SNPs (R^2)	0.997	0.948
Percentage of R^2 values ≥ 0.7	100	96.3

HWE, Hardy-Weinberg equilibrium; SNP, single nucleotide polymorphism. [a] See Supplementary Table 1. [b] See Supplementary Table 2. [c] Common was defined as minor allele frequency ≥ 0.03. [d] $p < 0.01$. [e] SNP dropping method by Weale *et al*. [14].

yses. We also used a Bayesian haplotype clustering method [18], with a fixed window size of six tagSNPs. Interestingly, the posterior distribution for the position of a possible disease mutation coincided with the region suggested by both fixed- and variable-sized haplotype analyses (Supplementary Figure 3 in Additional File 1).

Also, an analysis within groups of diabetes mellitus, menopausal hormone use or family history furthermore did not reveal any significant evidence for any tagSNP to be associated with breast cancer risk (data not shown).

EGF

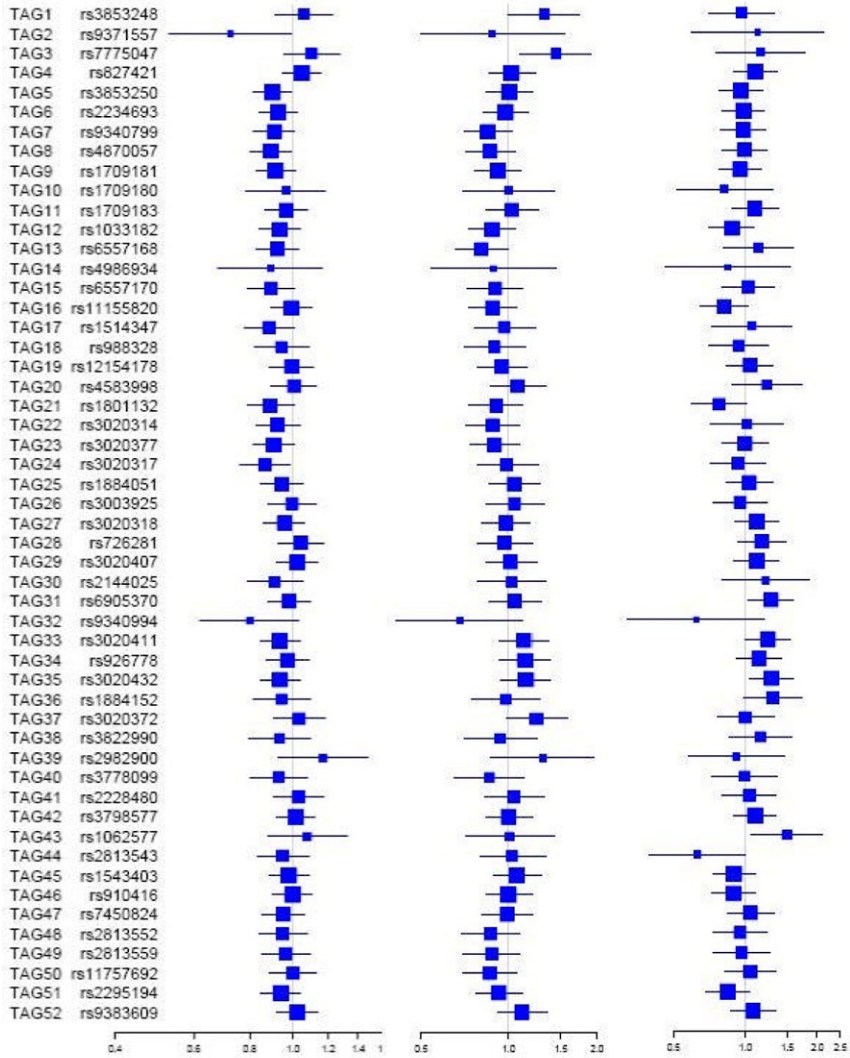
None of the tagSNPs in *EGF* showed association with breast cancer risk, NPI, or breast cancer survival that withstood multiple testing correction (Supplementary Figure 4 in Additional File 1). This lack of association was supported by the haplotype analysis.

Discussion

We had comprehensive SNP coverage of the entire *ESR1* and *EGF* genes and were thus able to study if there were any common variants in the genes that showed an association with breast cancer risk, NPI or breast cancer survival. No association was found between common variants in *ESR1* and NPI or breast cancer survival by single tagSNP analysis. A region between TAG26 (rs3003925) and TAG30 (rs2144025) in the *ESR1* gene showed a signal for association with breast cancer risk in the multi-locus analysis of five adjacent tagSNPs, but the result did not withstand multiple testing correction. Interestingly, the suggestive evidence from further haplotype analyses converge to this region, but its significance needs to be determined by follow-up analysis. None of the genotyped SNPs within this region were located in exons (all were in the middle of intron 4–5) and are thus unlikely to affect ESR1 protein structure. It is still a possibility however that the SNPs themselves, or one or more SNPs in LD with any of the SNPs, may effect the regulation of ESR1 protein expression. In fact, it has been shown that ESR1 protein overexpression is common in breast cancer [24]. Common variants within the *EGF* gene did not appear to affect the risk of developing breast cancer, developing a tumour with high NPI, or dying from the disease.

Our study was a well designed, population-based case-control study. Case ascertainment and case survival status were established using the nationwide, high-quality Cancer Registry and Causes of Death Registry in Sweden. Exposure status of the participants was determined using genotyping methods with low error rates from which all results underwent detailed quality control. We sought to obtain tissue samples from the deceased cases and those cases that had declined donation

Figure 1



Association of 52 tagging single nucleotide polymorphisms (tagSNPs) in *ESR1* with breast cancer risk, Nottingham Prognostic Index (NPI) and breast cancer survival. Left column: breast cancer risk. Middle column: NPI (case-only analysis). Right column: breast cancer survival. Squares and horizontal lines represent odds and hazard (survival analysis) ratios (change in risk with each addition of the rare allele) and their confidence intervals. Sizes of the squares reflect the minor allele frequencies. NPI was categorised into ≤ 4 or > 4.

of a blood sample, and were able to obtain the majority of the samples requested. The relative minor lack of tissue accessibility is unlikely to be related to our exposure, *ESR1* or *EGF* genetic variation, as it depended on the inability of the respective pathology department to retrieve the samples. The tissue sample availability was therefore random and could not have lead to selection bias. The main concern is that the non-participation of a small number of deceased cases might have reduced the power of our study, especially for the survival analysis. Furthermore, a problem might have arisen since we were not able to genotype seven tagSNPs in *ESR1* and one tagSNP in *EGF* in the tissue samples. If these eight tagSNPs were in fact associated with severe disease, the association with risk of breast cancer death might have been biased towards null in our study since we did not genotype all the severe cases. The fact that the results were not different when we restricted our analyses to the most severe cases among those who donated blood samples indicates that the eight tagSNPs were unlikely to be associated with severe disease.

In the selection stage of our study, we oversampled cases and controls that were long-term users of menopausal hormones and those that had self-reported diabetes mellitus. In the case of an association between the tagSNPs under study and menopausal hormone use or diabetes mellitus, this oversampling might have caused us to detect an artificial association between the tagSNPs and breast cancer risk. We therefore assessed if the tagSNPs were associated with menopausal

Table 3

Association between haplotypes reconstructed from *ESR1* TAGs 26–30 and breast cancer risk

		Haplotype proportions		
TAGs 26–30	Haplotypes	Cases ($n = 1,579$[a])	Controls ($n = 1,514$[a])	OR (95% CI)
Haplotype 1	ACAAC	0.57	0.56	1.00 (Reference)
Haplotype 2	GTGGT	0.09	0.10	0.91 (0.76–1.09)
Haplotype 3	ACAGC	0.06	0.06	1.05 (0.84–1.31)
Haplotype 4	ATGGT	0.06	0.06	0.89 (0.70–1.13)
Haplotype 5	ATGGC	0.06	0.05	1.05 (0.82–1.35)
Haplotype 6	GTGGC	0.05	0.03	1.46 (1.08–1.98)
Haplotype 7	GTAAC	0.03	0.04	0.72 (0.55–0.96)
	Rare[b]	0.08	0.08	0.94 (0.76–1.15)
Global p value[c]				0.0493

CI, confidence interval; SNP, single nucleotide polymorphism. [a] Information on at least one of the five tagSNPs. [b] Total of 18 rare haplotypes combined. Each haplotype has frequency below 3% among the controls. [c] Likelihood ratio test.

hormone use or diabetes mellitus. We found no connection between the factors and conclude that the oversampling is unlikely to have posed a problem in our study.

Most previous publications regarding the *ESR1* gene and breast cancer risk have included only a few polymorphisms in the gene. One study, however, genotyped 17 common SNPs in the *ESR1* gene and found three haplotypes to decrease breast cancer risk and one haplotype that increased the risk [6]. None of the haplotypes carried SNPs that were located in the region in *ESR1* we found to be associated with breast cancer risk. Two of the haplotypes that showed a protective effect against breast cancer risk (H4 and H6) carried our TAG21 (rs1801132, codon 325) [6]. We were not able to confirm this association using a window or a haplotype analysis.

The *Pvu*II (TAG6, rs2234693), *Xba*I (TAG7, rs9340799), codon 243 (TAG14, rs4986934) and codon 325 (TAG21, rs1801132) variants are among the most commonly studied polymorphisms in the *ESR1* gene. The first two have been suggested in a couple of studies to decrease the risk of endometrial cancer [25,26] and *Pvu*II might affect breast cancer survival depending on oestrogen receptor status of the tumour [27], but no consistent effect over studies has been shown for the four variants with regard to breast cancer risk [19,28-37]. We found no association between these SNPs and overall breast cancer risk, NPI or breast cancer survival.

Conclusion

We analysed common genetic variation in the *ESR1* and *EGF* genes in relation to breast cancer risk, tumour characteristics and breast cancer survival using a comprehensive haplotype tagging analysis. To our knowledge, this is the first systematic association study of these two genes for breast cancer susceptibility and prognosis. We located a region in *ESR1* which showed a moderate signal for association with breast cancer risk, but were unable to link common variation in the *EGF* gene with breast cancer aetiology or prognosis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KE, KC, SW, ETL, PH, KH and JL were involved in planning the study. YLL, YQL and CB administrated the genotyping analysis. KE, HD, YL, AS, KH and JL performed the statistical anal-

Table 4

Association of three haplotypes (TAGs 18–27) in *ESR1* with breast cancer risk, as implicated by the variable-sized sliding-window analysis

tagSNPs	Haplotype	Frequency in cases	Frequency in controls	OR (95% CI)	p Value[a]
TAG18–21	ACAC	0.0798	0.0595	1.39 (1.13–1.69)	0.0014
TAG18–24	ACAGCGC	0.0456	0.0562	0.70 (0.56–0.88)	0.0019
TAG18–27	GCAGCGCGGT	0.0697	0.0813	0.85 (0.70–1.03)	0.0933

CI, confidence interval; SNP, single nucleotide polymorphism. [a] One vs the others Chi-squared test.

ysis. KE, KH, PH and JL drafted the manuscript, and all the authors approved the manuscript.

Additional files

The following Additional files are available online:

Additional file 1

Additional file 1 is a Word file containing tables and figures with genotyping information, linkage disequilibrium maps and association analyses. See http://www.biomedcentral.com/content/supplementary/bcr1861-S1.doc

Acknowledgements

We would like to thank Christer Halldén at Swegene Laboratories, Malmö, Sweden, for overseeing the DNA isolation from the blood samples; Heng Khai Koon and Ong Eng Hu Jason for genotyping; Lim Siew Lan and Irene Chen for isolating DNA from paraffin-embedded tissue; Frans Verhoeff for processing the genotyping data; and Lena U Rosenberg, Mattias Hammarström and Eija Flygare for reviewing medical records. We are also grateful to all the women who took the time and effort to participate in this study and to Anna Christensson and Boel Bissmarck who obtained consent and coordinated the collecting of samples from the study participants. This study was supported by funding from the Agency for Science & Technology and Research of Singapore (A*STAR), the Susan G Komen Breast Cancer Foundation, the US National Institutes of Health (grant number R01 CA 104021) and the United States Army (Prime Award No. W81XWH-05-1-0314). KH acknowledges financial support from the Swedish Research Council.

References

1. Sherbet GV: **Hormonal influences on cancer progression and prognosis.** *Vitam Horm* 2005, **71:**147-200.
2. Persson I: **Estrogens in the causation of breast, endometrial and ovarian cancers – evidence and hypotheses from epidemiological findings.** *J Steroid Biochem Mol Biol* 2000, **74:**357-364.
3. Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, Tujague M, Strom A, Treuter E, Warner M, Gustafsson JÅ: **Estrogen receptors: how do they signal and what are their targets.** *Physiol Rev* 2007, **87:**905-931.
4. Couse JF, Lindzey J, Grandien K, Gustafsson JA, Korach KS: **Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse.** *Endocrinology* 1997, **138:**4613-4621.
5. Dickson RB, Lippman ME: **Growth regulation of normal and malignant breast epithelium.** In *The Breast: Comprehensive Management of Benign and Malignant Diseases* Edited by: Bland KI, Copeland EM. Philadelphia, PA: Saunders; 1998:518-536.
6. Gold B, Kalush F, Bergeron J, Scott K, Mitra N, Wilson K, Ellis N, Huang H, Chen M, Lippert R, Halldorsson BV, Woodworth B, White T, Clark AG, Parl FF, Broder S, Dean M, Offit K: **Estrogen receptor genotypes and haplotypes associated with breast cancer risk.** *Cancer Res* 2004, **64:**8891-8900.
7. Isola J, DeVries S, Chu L, Ghazvini S, Waldman F: **Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples.** *Am J Pathol* 1994, **145:**1301-1308.
8. **SNP** [http://www.ncbi.nlm.nih.gov/projects/SNP/]
9. **R Project** [http://www.r-project.org/]
10. Qin ZS, Niu T, Liu JS: **Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms.** *Am J Hum Genet* 2002, **71:**1242-1247.
11. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC: **Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study.** *Hum Hered* 2003, **55:**27-36.
12. Iles MM: **Obtaining unbiased estimates of tagging SNP performance.** *Ann Hum Genet* 2006, **70:**254-261.
13. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: **The structure of haplotype blocks in the human genome.** *Science* 2002, **296:**2225-2229.
14. Weale ME, Depondt C, Macdonald SJ, Smith A, Lai PS, Shorvon SD, Wood NW, Goldstein DB: **Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping.** *Am J Hum Genet* 2003, **73:**551-565.
15. Sundquist M, Thorstenson S, Brudin L, Nordenskjold B: **Applying the Nottingham Prognostic Index to a Swedish breast cancer population. South East Swedish Breast Cancer Study Group.** *Breast Cancer Res Treat* 1999, **53:**1-8.
16. Clayton D, Chapman J, Cooper J: **Use of unphased multilocus genotype data in indirect association studies.** *Genet Epidemiol* 2004, **27:**415-428.
17. Westfall P, Young S: *Resampling-based Multiple Testing: Examples and Methods for p Value Adjustment* New York: John Wiley & Sons; 1993.
18. Molitor J, Marjoram P, Thomas D: **Fine-scale mapping of disease genes with multiple mutations via spatial clustering techniques.** *Am J Hum Genet* 2003, **73:**1368-1384.
19. Wang J, Higuchi R, Modugno F, Li J, Umblas N, Lee J, Lui LY, Ziv E, Tice JA, Cummings SR, Rhees B: **Estrogen receptor alpha haplotypes and breast cancer risk in older Caucasian women.** *Breast Cancer Res Treat* 2007, **106:**273-280.
20. Chapman JM, Cooper JD, Todd JA, Clayton DG: **Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power.** *Hum Hered* 2003, **56:**18-31.
21. Gauderman WJ: **Sample size requirements for matched case-control studies of gene-environment interaction.** *Stat Med* 2002, **21:**35-50.
22. Manolio TA, Bailey-Wilson JE, Collins FS: **Genes, environment and the value of prospective cohort studies.** *Nat Rev Genet* 2006, **7:**812-820.
23. Einarsdottir K, Rosenberg LU, Humphreys K, Bonnard C, Palmgren J, Li Y, Chia KS, Liu ET, Hall P, Liu J, Wedrén S: **Comprehensive analysis of the ATM, CHEK2 and ERBB2 genes in relation to breast tumour characteristics and survival: a population-based case-control and follow-up study.** *Breast Cancer Res* 2006, **8:**R67.
24. Holst F, Stahl PR, Ruiz C, Hellwinkel O, Jehan Z, Wendland M, Lebeau A, Terracciano L, Al-Kuraya K, Janicke F, Sauter G, Simon R: **Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer.** *Nat Genet* 2007, **39:**655-660.
25. Iwamoto I, Fujino T, Douchi T, Nagata Y: **Association of estrogen receptor alpha and beta3-adrenergic receptor polymorphisms with endometrial cancer.** *Obstet Gynecol* 2003, **102:**506-511.
26. Weiderpass E, Persson I, Melhus H, Wedren S, Kindmark A, Baron JA: **Estrogen receptor alpha gene polymorphisms and endometrial cancer risk.** *Carcinogenesis* 2000, **21:**623-627.
27. Boyapati SM, Shu XO, Ruan ZX, Cai Q, Smith JR, Wen W, Gao YT, Zheng W: **Polymorphisms in ER-alpha gene interact with estrogen receptor status in breast cancer survival.** *Clin Cancer Res* 2005, **11:**1093-1098.
28. Shen Y, Li DK, Wu J, Zhang Z, Gao E: **Joint effects of the CYP1A1 MspI, ERalpha *Pvu* II, and ERalpha *Xba*I polymorphisms on the risk of breast cancer: results from a population-based case-control study in Shanghai, China.** *Cancer Epidemiol Biomarkers Prev* 2006, **15:**342-347.
29. Slattery ML, Sweeney C, Herrick J, Wolff R, Baumgartner K, Giuliano A, Byers T: **ESR1, AR, body size, and breast cancer risk in Hispanic and non-Hispanic white women living in the Southwestern United States.** *Breast Cancer Res Treat* 2007, **105:**327-335.

30. Onland-Moret NC, van Gils CH, Roest M, Grobbee DE, Peeters PH: **The estrogen receptor alpha gene and breast cancer risk (The Netherlands).** *Cancer Causes Control* 2005, **16:**1195-1202.
31. Cai Q, Shu XO, Jin F, Dai Q, Wen W, Cheng JR, Gao YT, Zheng W: **Genetic polymorphisms in the estrogen receptor alpha gene and risk of breast cancer: results from the Shanghai Breast Cancer Study.** *Cancer Epidemiol Biomarkers Prev* 2003, **12:**853-859.
32. Shin A, Kang D, Nishio H, Lee MJ, Park SK, Kim SU, Noh DY, Choe KJ, Ahn SH, Hirvonen A, Kim JH, Yoo K-Y: **Estrogen receptor alpha gene polymorphisms and breast cancer risk.** *Breast Cancer Res Treat* 2003, **80:**127-131.
33. Modugno F, Zmuda JM, Potter D, Cai C, Ziv E, Cummings SR, Stone KL, Morin PA, Greene D, Cauley JA: **Association of estrogen receptor alpha polymorphisms with breast cancer risk in older Caucasian women.** *Int J Cancer* 2005, **116:**984-991.
34. Curran JE, Lea RA, Rutherford S, Weinstein SR, Griffiths LR: **Association of estrogen receptor and glucocorticoid receptor gene polymorphisms with sporadic breast cancer.** *Int J Cancer* 2001, **95:**271-275.
35. Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF: **A systematic review of genetic polymorphisms and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 1999, **8:**843-854.
36. Schubert EL, Lee MK, Newman B, King MC: **Single nucleotide polymorphisms (SNPs) in the estrogen receptor gene and breast cancer susceptibility.** *J Steroid Biochem Mol Biol* 1999, **71:**21-27.
37. Wedren S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvanen AC, Kindmark A, Landegren U, Fermer ML, Stiger F, Persson I, Baron J, Weiderpass E: **Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study.** *Breast Cancer Res* 2004, **6:**R437-449.

Hypothesis/Commentary

Plasma Volume Expansion in Pregnancy: Implications for Biomarkers in Population Studies

Jessica M. Faupel-Badger,[1,2] Chung-Cheng Hsieh,[5] Rebecca Troisi,[3] Pagona Lagiou,[6,7] and Nancy Potischman[4]

[1]Division of Cancer Prevention, [2]Mammary Biology and Tumorigenesis Laboratory, Center for Cancer Research, [3]Division of Cancer Epidemiology and Genetics, and [4]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland; [5]University of Massachusetts Medical School, Worcester, Massachusetts; [6]Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece; and [7]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts

Abstract

There is a growing body of literature focused on endogenous hormone exposures during pregnancy and subsequent cancer risk for both mother and offspring. Examples of these studies include those focused on the biological mechanism for the association of preeclampsia with reduced risk of breast cancer for mother and female offspring or studies that have examined hormone concentrations during pregnancy between different ethnic groups who vary in their rates of breast cancer incidence. Although these studies seem relatively straightforward in conception and analysis, measurement of the concentration of hormones and other biomarkers in pregnant subjects is influenced by plasma volume expansion (PVE). During pregnancy, the maternal plasma volume expands 45% on average to provide for the greater circulatory needs of the maternal organs. Consequently, serum protein and hormone concentrations are greatly altered when comparing the pregnant with nonpregnant state. Assessing PVE also is complicated by the vast individual variation in PVE, ranging from minimal to a 2-fold increase. We propose that PVE needs to be evaluated when comparing biomarker concentrations during pregnancy in two populations that may differ with respect to PVE. Small body size is associated with lower PVE compared with higher body size. Therefore, we hypothesize that variation in PVE will influence the interpretation of differences in biomarker concentrations across population groups with respect to the etiologic significance of the biomarker to the disease under study (e.g., breast cancer). It is possible that some observations may be due only to differences in dilution between the two groups. We present PVE as a topic for consideration in population-based studies, examples of the types of studies where PVE may be relevant, and our own analysis of one such study in the text below. (Cancer Epidemiol Biomarkers Prev 2007;16(9):1720–3)

Introduction to Plasma Volume Expansion in Pregnancy

During pregnancy, maternal plasma volume increases to meet the greater circulatory needs of the placenta and maternal organs (e.g., uterus, breasts, skin, and kidneys), with an average increase of ~45% (1-5). There are vast differences among women, however, from a minimal change to a doubling in plasma volume (1, 6, 7). Several factors can influence plasma volume expansion (PVE) including maternal pre-pregnancy body mass index (BMI; ref. 8). In studies that have examined ethnic differences in PVE, populations that are on average shorter and weigh less exhibit less absolute change in plasma volume (9-11). This was well documented in a report comparing plasma volumes between Indians and Europeans (10). Furthermore, of studies examining PVE across different ethnicities, all show vast individual variation in PVE as well (9-11). Much of the variation is unexplained, although it is known that PVE is positively associated with parity (12), multiple births (13-15), higher birth weight (16, 17), and increased maternal pre-pregnancy BMI (8) and inversely associated with conditions of decreased fetal growth (e.g., intrauterine growth restriction) and compromised placental development (e.g., preeclampsia; refs. 1, 2, 18-20).

In clinical practice, PVE or 'hemodilution' is addressed by modifying the criteria for diagnostic biomarkers. For example, the cut-point for anemia or iron deficiency, which is defined by hemoglobin <12.0 g/dL in the

Received 4/3/07; revised 5/29/07; accepted 6/22/07.

Grant support: Division of Cancer Prevention and the Center for Cancer Research, National Cancer Institute (J.M. Faupel-Badger) and W81XWH-05-1-0314 Innovator Award and U.S. Department of Defense Breast Cancer Research Program, Office of the Congressionally Directed Medical Research Programs (P. Lagiou and C-C. Hsieh). J.M. Faupel-Badger is a fellow in the National Cancer Institute Cancer Prevention Fellowship Program.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Jessica M. Faupel-Badger, Mammary Biology and Tumorigenesis Laboratory, Center for Cancer Research, National Cancer Institute, MSC 4254, 37 Convent Drive, Bethesda, MD 20892. Phone: 301-443-3076; Fax: 301-402-0711. E-mail: badgerje@mail.nih.gov

doi:10.1158/1055-9965.EPI-07-0311

nonpregnant state, is changed to <11.0 g/dL in the first and third trimesters and <10.5 g/dL in the second trimester. This corresponds to the increase in plasma volume starting at 6 to 10 weeks of gestation that rises sharply through the second trimester, before beginning to plateau at 32 weeks (1).

PVE and Population Studies

Highlighting diagnosis of anemia in clinical practice shows that PVE can have significant effects on biomarker concentrations. The implications of interindividual and between-group variability in PVE, however, have generally not been addressed in large population-based research studies involving pregnant subjects. In etiologic studies, this variability could introduce bias in the form of confounding if it were related to both the factor under study (e.g., preeclampsia) and the biomarker. Thus, accounting for PVE would be essential for a study focused on elucidating biomarkers involved in the causal pathway of a disease. For example, in preeclampsia, there is less PVE than in uncomplicated pregnancies (1, 2, 19). High concentrations of a biomarker such as sex hormone binding globulin (21) or soluble fms-like tyrosine kinase (1, 22, 23) among preeclamptic women compared with women who have uncomplicated pregnancies may reflect an etiologically relevant difference and/or the lower hemodilution in preeclampsia. We hypothesize that variation in PVE will affect the interpretation of differences in biomarker concentrations between individuals or population groups, especially with respect to etiologic significance, and that evaluation of PVE should be considered in population studies involving pregnant women.

Other examples in which individual variation in PVE may affect interpretation include when biomarkers are used to measure the success of nutritional intervention during pregnancy (24). In some areas of Asia, it is not uncommon for pregnant women to be deficient in multiple micronutrients (24, 25). To address this issue, one study among rural pregnant women in Nepal evaluated the effects of micronutrient supplementation on serum retinol, folate, riboflavin, and 25-hydroxyvitamin D concentrations (26). Assuming this population has wide variation in individual PVE, as has been shown in other populations, PVE may affect the perceived success of this intervention in individual women. Furthermore, variability in nutritional status and hydration across individuals could influence the apparent concentrations of nutrients of interest in this study via effects on PVE.

Measurement of PVE, Biomarker Concentrations, and Relation to Outcomes

PVE also may be an important factor in studies focused on understanding what biological features mediate associations of pregnancy characteristics with subsequent development of chronic disease in either the mother or offspring. For example, preeclampsia is associated with decreased breast cancer risk for both the mother and female offspring (27). Population-based studies of circulating biomarkers in preeclamptic and normotensive pregnancies have been conducted to elucidate possible biological mechanisms. Serum androgen concentrations at delivery are observed to be higher in preeclamptic women than in those with normal pregnancies (28). Progesterone concentrations also seem higher in maternal serum during the 27th week of pregnancy when comparing preeclamptic with normal pregnancies (29). Could the lower PVE that preeclamptic women experience relative to normotensive women explain their apparently higher concentration of serum hormones? And if so, are differences due to PVE of etiologic importance if the target tissues are exposed to the same total amount of the hormones?

Direct methods of measuring plasma volume are labor intensive and not easily adapted to large studies. Methods to approximate plasma volume involve labeling of albumin, with Evan's blue dye being the most common technique (11, 30). This method requires sampling plasma after injecting the patient with Evan's blue dye and allowing time (a minimum of 10 min) for sufficient mixing of the dye with plasma. The decay of Evan's blue dye is measured by the absorbance at 610 nm and plasma volume in milliliters per kilogram can be calculated from the absorbance measurement (30). In addition to patient monitoring during the dye injection and sample collection, it also is recommended that the patient observe a 30-min rest period before injection and that women in late pregnancy lay on their sides during injection to promote mixing of the dye (11, 30). Measuring Evan's blue dye at one point in time can approximate a measurement of plasma volume; however, to measure the increase in plasma volume, an individual would need to have the Evan's blue dye measurement completed more than once during pregnancy. These measurements across time points would then be used to determine the amount of PVE for the individual. This technique involves significant patient burden and is not practical for large studies.

In population-based studies, accounting for PVE in the data analysis by adjustment for factors highly predictive of PVE is more feasible. We reanalyzed data from a published study comparing maternal hormone concentrations in women from Boston and Shanghai (31) to determine if adjusting for correlates of PVE would affect the observed differences in estradiol concentrations. The women for this study were recruited from urban clinics affiliated with Beth Israel Hospital in Boston and three urban and one rural clinic in China affiliated with Shanghai Medical University. The women were all under the age of 40 years and were either Caucasian in Boston or Chinese in Shanghai. The purpose of the published study was to explore the hypothesis that fetal exposures, such as high *in utero* hormone concentrations, may be associated with the development of breast cancer in the offspring, using two populations with different disease incidence (31). Two serum samples, collected at 16 and 27 weeks of gestation, were used to evaluate estradiol, estriol, prolactin, progesterone, growth hormone, albumin, and sex hormone binding globulin. Given that women in Shanghai have approximately one fifth the incidence of breast cancer of women in Boston (32), it was hypothesized that the women from Shanghai would have lower hormone concentrations. In contrast, for every compound measured, the women from Shanghai had significantly higher concentrations with the exception of progesterone at 27 weeks. The results from this

Table 1. Percentage difference in maternal estradiol at 16 wks of gestation comparing Chinese with U.S. concentrations by strata of pre-pregnancy BMI, uniparous subjects only

	Adjusted for:			With further adjustment for:				
Pre-pregnancy BMI (kg/m^2)	Boston (n)	Shanghai (n)	Maternal age, gestational age	Albumin	Albumin, pre-weight	Albumin, height	Albumin, pre-BMI	Albumin, birth weight
<19.1	21	84	37.3*	39.5*	38.2*	35.2*	40.5*	40.8*
19.1-20.5	30	68	24.5*	31.0*	19.8	20.5	30.8*	31.0*
20.6-22.2	34	46	12.6	16.4	15.9	16.6	16.7	14.7
22.3+	36	22	34.8*	35.6*	34.3*	37.7*	34.3*	35.8*
All subjects	121	200	26.7*	30.5*	27.7*	29.0*	29.4*	30.6*

NOTE: Adjusted for maternal age, pre-pregnancy weight, pre-pregnancy BMI, height, and albumin measured at week 16, and offspring birthweight and gestational age.
*$P < 0.05$.

study were unexpected as Chinese women are known to have lower circulating estradiol concentrations than Caucasian women in the nonpregnant state (33-35).

We hypothesized that differences in PVE between the Chinese and U.S. populations could explain the higher estradiol concentrations observed among the Chinese. Asian women have lower values for several correlates of PVE, such as height, weight, and infant birth weight, when compared with Caucasian women (31, 36, 37). To assess whether the higher hormone concentrations present in Asian women were a result of less PVE compared with Caucasian women, we analyzed the data set from Lipworth et al. (31) adjusting for or stratifying on correlates of PVE. The analysis was restricted to uniparous (i.e., first pregnancy) women because parity is known to affect PVE (11). Regression analysis with log-transformed estradiol as the dependent variable and with adjustment for maternal age, pre-pregnancy weight, height, pre-pregnancy BMI, birth weight, and albumin concentration allowed calculation of percentage difference in estradiol concentrations between the Shanghai and Boston populations for subjects with complete data on covariates (131 subjects for Boston and 220 in Shanghai). Similar to results in the original study, among all uniparous women in our analysis, the estradiol concentrations in the women from Shanghai were 26.7% higher than the estradiol concentrations in the women from Boston, at 16 weeks of gestation, after adjusting for maternal age and gestational week. Table 1 presents the percentage difference among women at 16 weeks of gestation within the same category of pre-pregnancy BMI to attempt to further account for PVE. The differences in estradiol concentrations between the Boston and Shanghai populations were strongest at 16 weeks and adjustment for correlates of PVE, including maternal pre-pregnancy BMI and birth weight, had no effect. Results for analyses of estradiol concentrations at week 16 were similar when all subjects from the Lipworth study, uniparous and multiparous, were included in the analyses (data not shown).

Conclusions and Comments for Future Analysis of PVE

Our reanalysis of the estradiol data shows that the difference in estradiol concentrations between the Boston and Shanghai populations cannot be explained by adjusting for currently accepted correlates of PVE. Therefore, we have two possible conclusions. The difference in estradiol concentrations between these two populations may be real and due to factors other than differences in PVE. Alternatively, it is possible that pre-pregnancy BMI may not be a sufficient marker of PVE. In nonpregnant women, it may be possible to adjust for differences in plasma volume by including BMI in the model, as BMI is highly correlated with plasma volume in healthy, nonpregnant individuals (8). Whereas pre-pregnancy BMI is currently an accepted correlate for PVE, other factors (e.g., physiologic processes related to adaptation to pregnancy and size of fetus) besides maternal size can affect PVE, thus suggesting that pre-pregnancy BMI may not be the most adequate correlate of PVE, although it may be the best one in use at this time.

Although reanalysis of the data from the study by Lipworth et al. did not support our hypothesis that pre-pregnancy BMI would be a sufficient predictor to account for PVE, it may be worth considering this type of analysis in other studies with larger numbers or among different ethnicities. In addition, to fully account for any effects of PVE, it may be that new proxies or methods to evaluate PVE need to be developed for practical use in large studies. While new proxies or methods are being explored, other avenues to measure hormone exposure during pregnancy could be used to circumvent the issue of PVE. For example, studies focused on *in utero* hormone exposure and subsequent disease risk in the offspring could also examine cord blood samples for the measurement of hormones or other biomarkers (38). These samples would be insulated from variations in dilution present in the maternal circulation and may be more relevant to fetal exposure than conclusions based on maternal serum. In addition to cord blood, saliva or urine may also provide a source for hormone measurements that may be less influenced by PVE. Future studies with samples to measure the correlation of maternal serum hormone concentrations to concentrations in cord blood, urine and saliva would be helpful to determine if the magnitude increase in hormone concentrations seen in serum is reflected in the alternative method. It would also be necessary to evaluate whether the alternative method may also be influenced by PVE. Another approach would be to evaluate changes in a biomarker over the course of a pregnancy, perhaps elucidating different patterns in those that develop a pregnancy condition from those who remain normal.

In conclusion, given the vast individual and group differences in PVE, its implications for the interpretation of results from biomarkers studies should be considered. We hope this commentary begins a dialogue about the implications of PVE and its evaluation for the types of studies outlined above. Further evaluation of PVE is necessary to determine its potential effect on the interpretation of observed results with regard to etiology and the degree to which it needs to be addressed in study design and/or analysis.

References

1. Blackburn ST, Loper DL. Maternal, fetal, and neonatal physiology. A clinical perspective. Philadelphia: W.B. Saunders Co.; 1992. p. 160–2, 171, 202–3, 222–8.
2. Brinkman CR. Maternal cardiovascular and renal disorders. Maternal-fetal medicine. Philadelphia: W.B. Saunders; 1984. p. 679–90.
3. Chesley LC. Plasma and red cell volumes during pregnancy. Am J Obstet Gynecol 1972;112:440.
4. Peck TM, Arais F. Hematologic changes associated with pregnancy. Clin Obstet Gynecol 1979;22:785.
5. Pritchard JA. Changes in blood volume during pregnancy and delivery. Anesthesiology 1965;26:393.
6. Pritchard JA, MacDonald PC, Gant N. Williams' obstetrics. New York: Appleton-Century-Crofts; 1985. p. 191–4.
7. Anderson HM. Maternal hematologic disorders. Maternal-fetal medicine. Philadelphia: W.B. Saunders; 1984. p. 795–832.
8. Dagher FJ, Lyons JH, Finlayson DC, Shamsai J, Moore FD. Blood volume measurement: a critical study prediction of normal values: controlled measurement of sequential changes: choice of a bedside method. Adv Surg 1965;1:69–109.
9. Harrison KA. Blood volume changes in normal pregnant Nigerian women. J Obstet Gynaecol Br Commonw 1966;73:717–23.
10. Hutchins CJ. Plasma volume changes in pregnancy in Indian and European primigravidae. Br J Obstet Gynaecol 1980;87:586–9.
11. Hytten F. Blood volume changes in normal pregnancy. Clin Haematol 1985;14:601–12.
12. Campbell DM, MacGillivray I. Comparison of maternal response in first and second pregnancies in relation to baby weight. J Obstet Gynaecol Br Commonw 1972;79:684–93.
13. Fullerton WT, Hytten FE, Klopper AI, McKay E. A case of quadruplet pregnancy. J Obstet Gynaecol Br Commonw 1965;72:791–6.
14. Letsky E. The haematological system. In: Hytten FE, Chamberlain GVP, editors. Clinical physiology in obstetrics. Oxford: Blackwell Scientific Publications; 1980. p. 43–78.
15. Rovinsky JJ, Jaffin H. Cardiovascular hemodynamics in pregnancy. I. Blood and plasma volumes in multiple pregnancy. Am J Obstet Gynecol 1965;93:1–15.
16. Hytten FE, Paintin DB. Increase in plasma volume during normal pregnancy. J Obstet Gynaecol Br Emp 1963;70:402–7.
17. Pirani BB, Campbell DM, MacGillivray I. Plasma volume in normal first pregnancy. J Obstet Gynaecol Br Commonw 1973;80: 884–7.
18. Bentley DP. Iron metabolism and anaemia in pregnancy. Clin Haematol 1985;14:613.
19. Salas SP, Marshall G, Gutierrez BL, Rosso P. Time course of maternal plasma volume and hormonal changes in women with preeclampsia or fetal growth restriction. Hypertension 2006;47:203–8.
20. Salas SP, Rosso P, Espinoza R, et al. Maternal plasma volume expansion and hormonal changes in women with idiopathic fetal growth retardation. Obstet Gynecol 1993;81:1029–33.
21. Wolf M, Sandler L, Munoz K, et al. First trimester insulin resistance and subsequent preeclampsia: a prospective study. J Clin Endocrinol Metab 2002;87:1563–8.
22. Levine RJ, Lam C, Qian C, et al. Soluble endoglin and other circulating antiangiogenic factors in preeclampsia. N Engl J Med 2006;355:992–1005.
23. Levine RJ, Maynard SE, Qian C, et al. Circulating angiogenic factors and the risk of preeclampsia. N Engl J Med 2004;350:672–83.
24. Ahmed F, Mahmuda I, Sattar A, Akhtaruzzaman M. Anaemia and vitamin A deficiency in poor urban pregnant women of Bangladesh. Asia Pac J Clin Nutr 2003;12:460–6.
25. Jiang T, Christian P, Khatry SK, Wu L, West KP, Jr. Micronutrient deficiencies in early pregnancy are common, concurrent, and vary by season among rural Nepali pregnant women. J Nutr 2005;135:1106–12.
26. Christian P, Jiang T, Khatry SK, et al. Antenatal supplementation with micronutrients and biochemical indicators of status and subclinical infection in rural Nepal. Am J Clin Nutr 2006;83:788–94.
27. Innes KE, Byers TE. Preeclampsia and breast cancer risk. Epidemiology 1999;10:722–32.
28. Troisi R, Potischman N, Roberts JM, et al. Maternal serum oestrogen and androgen concentrations in preeclamptic and uncomplicated pregnancies. Int J Epidemiol 2003;32:455–60.
29. Tamimi R, Lagiou P, Vatten LJ, et al. Pregnancy hormones, preeclampsia, and implications for breast cancer risk in the offspring. Cancer Epidemiol Biomarkers Prev 2003;12:647–50.
30. el-Sayed H, Goodall SR, Hainsworth R. Re-evaluation of Evans blue dye dilution method of plasma volume measurement. Clin Lab Haematol 1995;17:189–94.
31. Lipworth L, Hsieh CC, Wide L, et al. Maternal pregnancy hormone levels in an area with a high incidence (Boston, USA) and in an area with a low incidence (Shanghai, China) of breast cancer. Br J Cancer 1999;79:7–12.
32. Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2002: Cancer incidence, mortality and prevalence Worldwide. In: IARC, editor. CancerBase no. 5. version 2.0 2006. Lyon: IARC Press; 2004.
33. Bernstein L, Yuan JM, Ross RK, et al. Serum hormone levels in premenopausal Chinese women in Shanghai and white women in Los Angeles: results from two breast cancer case-control studies. Cancer Causes Control 1990;1:51–8.
34. Key TJ, Chen J, Wang DY, Pike MC, Boreham J. Sex hormones in women in rural China and in Britain. Br J Cancer 1990;62:631–6.
35. Ursin G, Wilson M, Henderson BE, et al. Do urinary estrogen metabolites reflect the differences in breast cancer risk between Singapore Chinese and United States African-American and white women? Cancer Res 2001;61:3326–9.
36. Lagiou P, Hsieh CC, Trichopoulos D, et al. Birthweight differences between USA and China and their relevance to breast cancer aetiology. Int J Epidemiol 2003;32:193–8.
37. Wen SW, Kramer MS, Usher RH. Comparison of birth weight distributions between Chinese and Caucasian infants. Am J Epidemiol 1995;141:1177–87.
38. Troisi R, Potischman N, Roberts JM, et al. Correlation of serum hormone concentrations in maternal and umbilical cord samples. Cancer Epidemiol Biomarkers Prev 2003;12:452–6.

British Journal of Cancer (2008) 98, 660 – 663

www.bjcancer.com

Correlation of umbilical cord blood haematopoietic stem and progenitor cell levels with birth weight: implications for a prenatal influence on cancer risk

WC Strohsnitter[1], TM Savarese[2,*], HP Low[3], DP Chelmow[1], P Lagiou[4], M Lambe[5], K Edmiston[6], Q Liu[2], I Baik[2], KL Noller[1], H-O Adami[5,7], D Trichopoulos[7] and C-C Hsieh[*,2,5,7]

[1]Department of Obstetrics and Gynecology, Tufts-New England Medical Center, Boston, MA, USA; [2]Department of Cancer Biology and Cancer Center, University of Massachusetts Medical School, Worcester, MA, USA; [3]Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA; [4]Department of Hygiene and Epidemiology, School of Medicine, Athens University, Athens, Greece; [5]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [6]Division of Hematology/Oncology, Department of Medicine, UMass Medical School and UMass Memorial Health Care, Worcester, MA, USA; [7]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

We examined the relation with birth weight and umbilical cord blood concentrations of haematopoietic stem and progenitor populations in 288 singleton infants. Across the whole range of birth weight, there was a positive relation between birth weight and $CD34^+CD38^-$ cells, with each 500 g increase in birth weight being associated with a 15.5% higher (95% confidence interval: 1.6–31.3%) cell concentration. $CD34^+$ and $CD34^+$ c-*kit*$^+$ cells had J-shaped relations and CFU-GM cells had a U-shaped relation with birth weight. Among newborns with $\geqslant$3000 g birth weights, concentrations of these cells increased with birth weight, while those below 3000 g had higher stem cell concentrations than the reference category of 3000–3499 g. Adjustment for cord blood plasma insulin-like growth factor-1 levels weakened the stem and progenitor cell–birth weight associations. The positive associations between birth weight and stem cell measurements for term newborns with a normal-to-high birth weight support the stem cell burden hypothesis of cancer risk.

British Journal of Cancer (2008) **98,** 660–663. doi:10.1038/sj.bjc.6604183 www.bjcancer.com
Published online 5 February 2008

Keywords: birth weight; cancer risk; prenatal exposure; stem cell

The *in utero* environment and perinatal factors may influence cancer risk of the offspring later in life (Trichopoulos, 1990). One parameter that reflects *in utero*/perinatal influences, that is birth weight, has been positively correlated with subsequent risk of childhood cancer (Schüz and Forman, 2007) and, in adults, breast (Michels and Xue, 2006), prostate (Eriksson *et al*, 2007) and colorectal cancers (Nilsen *et al*, 2005) and, indeed, overall cancer risk (Ahlgren *et al*, 2007). Mechanistically, a 'stem cell burden' theory (Adami *et al*, 1995) has been proposed to account for the positive relationship between birth weight and the risk of certain cancers, especially that of the breast. By this hypothesis, the levels of *in utero*/perinatal mitogens and other factors determine the size of the stem cell pools in the developing fetus; elevated tissue stem cell numbers drive the formation of larger organs and hence might be associated with larger birth weights. The greater the stem cell pool size, however, the greater the chance that one of the stem cells will be mutated by a carcinogen, or undergo a DNA replicative error, initiating oncogenic transformation. Hence, individuals with high birth weights might be at greater lifetime cancer risk (Trichopoulos *et al*, 2005).

The stem cell burden theory predicts (1) that the *in utero* levels of particular mitogens should correlate positively with stem cell population levels and (2) that the stem cell levels should correlate positively with birth weight. In a previous study, we demonstrated that the umbilical cord blood concentrations of various haematopoietic stem and progenitor populations correlated with cord blood plasma levels of particular mitogens, especially insulin-like growth factor-1 (IGF-1) (Savarese *et al*, 2007). Here, we determine whether or not these measurable haematopoietic stem cell/progenitor values, serving as surrogates of overall stem cell potential, are positively associated with birth weight.

MATERIALS AND METHODS

The umbilical cord blood study protocol was approved by the institutional review boards of the American Red Cross, the University of Massachusetts Medical School, the University of Massachusetts/Memorial Health Care System, St. Vincent's

*Correspondence: Dr C-C Hsieh, Division of Biostatistics and Epidemiology, Department of Cancer Biology, UMass Cancer Center, University of Massachusetts Medical School, Worcester, MA, USA;
E-mail: chung.hsieh@umassmed.edu
* Deceased.
Received 29 October 2007; revised 30 November 2007; accepted 7 December 2007; published online 5 February 2008

Hospital and Tufts-New England Medical Center (T-NEMC). Consenting study subjects were recruited from one of two sources: (1) participants in the Worcester, MA-based American Red Cross cord blood program (ACBP), in which a haematopoietic stem cells from umbilical cord blood were collected for possible transplantation, from August 2002 to June 2003, and (2) pregnant women delivering at T-NEMC from October 2004 to April 2006. All the cord blood samples were from full-term (gestational age $\geqslant 37$ weeks) singleton infants. The processing of samples, which includes the determination of cord blood volumes, the determination of initial levels of total nucleated cells (TNC) and mononuclear cells (MNC) before centrifugations or manipulation, the quantitation of haematopoietic stem/progenitor cell populations and the determination of cord blood plasma hormone levels, have been described (Savarese *et al*, 2007). The haematopoietic stem and progenitor populations that were quantitated (1) $CD34^+$ cells, a heterogeneous population of early multipotent stem and progenitor cells, committed progenitors and differentiating cells (Xiao and Dooley, 2000); (2) $CD34^+CD38^-$ cells, which represent more primitive stem cells depleted of lineage-committed precursors (Xiao and Dooley, 2000); (3) $CD34^+$c-*kit*$^+$ cells, which also represent a more primitive stem cell population that has relatively high cloning efficiencies in semisolid culture (Sharkey *et al*, 1994; Mayani and Lansdorp, 1998); and (4) granulocyte–macrophage colony forming units (CFU-GM), a functional measure of the number of proliferative granulocyte/macrophage-committed haematopoietic precursor cells (Abboud *et al*, 1992; Hoffbrand *et al*, 2001).

Birth weight was first studied as a categorical variable (<3000, 3000–3499, 3500–3999 and $\geqslant 4000$ g). Geometric means for the stem cell measurements were estimated within the indicated categories of birth weight. Multivariate linear regression was used to examine the association between natural log-transformed measures of stem cell potential (dependent variable) and birth weight (independent variable, using 3000–3499 g as the reference), adjusting for maternal and neonatal characteristics (mother's age, race of parents, number of live births, gestation duration, baby's gender, delivery time and study site). To assess whether there was an underlying linear trend, birth weight was next analyzed as a continuous variable across the whole range of birth weight values with the effect estimates expressed for each 500 g increase in birth weight. The fitted coefficients from the regression analyses were exponentiated to obtain the estimated proportional change in birth weight associated with each independent variable. Statistical significance was set at 0.05 (two-sided). Levels of IGF-1, which had the strongest association with levels of stem cells among the hormones and growth factors examined in a previous analysis (Savarese *et al*, 2007), were further adjusted to explore its influence on the association between birth weight and stem cell measurements.

RESULTS

The characteristics of the study subjects are shown in Table 1. Subjects from the ACBP and T-NEMC patient groups had similar age and gestation duration. Parental ethnicity was more varied in the T-NEMC samples, while the ACBP subjects had higher parities, more male newborns and lower birth weights.

The associations were analysed in multivariate analysis adjusting for maternal age, parental race, parity, gestation duration, gender, delivery time and study site (Table 2, upper panel). There was a J-shaped association between birth weight categories and concentrations of TNC (lymphocytes, monocytes and granulocytes), as well as a J-shaped relation with MNC (lymphocytes and monocytes), both including more differentiated cells. Among the stem cell populations, there was a positive association with $CD34^+CD38^-$ cells across the whole range of birth weight categories, with each 500 g increase being associated with 15.5% higher levels of this cell population (95% confidence interval: 1.6, 31.3%). A J-shaped relation was observed for the $CD34^+$ and $CD34^+$c-*kit*$^+$ cells: for birth weights of 3000 g or greater, stem cell concentrations increased with birth weight, while the lowest category of <3000 g had higher levels than the category of 3000–3499 g. For CFU-GM, an approximate U-shaped relation was observed, with the lowest birth weight category having the highest levels of this cell population (Table 2, upper panel).

Table 1 Characteristics of the study subjects by study site

Characteristics	ACBP ($n=39$) Mean ± s.d. or *N* (%)	T-NEMC ($n=249$) Mean ± s.d. or *N* (%)
Mother's age (years)	29.7 ± 5.1	30.1 ± 5.6
Parity		
1	14 (35.9)	104 (45.0)
2	12 (30.8)	64 (27.7)
3	7 (18.0)	39 (16.9)
4 or more	6 (15.4)	24 (10.4)
Race/ethnicity of mother and biological father		
Both Caucasian	36 (92.3)	111 (52.4)
Both African-American	1 (2.6)	19 (9.0)
Both Asian	0 (0.0)	39 (18.4)
Both Hispanic	0 (0.0)	14 (6.6)
Mixed	2 (5.1)	29 (13.7)
Gestation duration (weeks)	39.6 ± 1.4	39.6 ± 1.2
Newborn gender		
Male	21 (55.3)	113 (50.0)
Female	17 (44.7)	113 (50.0)
Birth weight (g)	3313.1 ± 450.6	3416.4 ± 428.2
<3000	11 (28.2)	40 (16.1)
3000–3499	15 (38.5)	110 (44.2)
3500–3999	11 (28.2)	77 (30.9)
$\geqslant 4000$	2 (5.1)	22 (8.8)

Abbreviations: ACBP = American Red Cross Cord Blood Program; T-NEMC = Tufts-New England Medical Center. Data on some variables were unavailable for subjects with missing values.

Adjusting for cord blood plasma levels of IGF-1 in the multivariate analysis weakens the association (Table 2, lower panel). The association with $CD34^+CD38^-$ remained positive but was no longer statistically significant: each 500 g increase in birth weight was associated with a 7.9% increase in this cell subpopulation (95% confidence interval: −6.2, 24.0%). The lowest weight category continued to have the highest CFU-GM cells after adjusting for IGF-1 levels.

We conducted further analyses adjusting for other hormones and in samples from different ethnic groups. Adjusting for estriol or insulin-like growth factor binding protein-3, which had statistically significant but weaker associations with cord blood levels of stem cells (Savarese *et al*, 2007), in place of IGF-1, had much less effect on the association with birth weight (data not shown). A linear relation between stem cell measurements and birth weight was observed for newborns whose parents were Caucasian, but did not have a consistent shape among samples in the mixed non-Caucasian group (Table 3).

DISCUSSION

The stem cell burden hypothesis has been invoked as an explanation for the positive link between birth weight and risk for both childhood and adult cancers (Adami *et al*, 1995). This hypothesis proposes that *in utero* environments that promote

Table 2 Multiple linear regression analysis for the association between measurements of haematopoietic stem cell populations and birth weight

		Birth weight (g) in categories				Birth weight per 500 g
Analytic model	**Cell measurements**	**<3000 % Difference (95% CI)**	**3000–3499 Reference (geometric mean[a])**	**3500–3999 % Difference (95% CI)**	**⩾4000 % Difference (95% CI)**	**% Difference (95% CI)**
Adjusted for core covariates[b]	TNC	2.1 (−8.5, 14.0)	0.0 (15.00)	7.8 (−1.7, 18.3)	8.4 (−8.1, 27.8)	3.0 (−2.5, 8.9)
	MNC	5.7 (−6.4, 19.3)	0.0 (6.91)	6.5 (−3.9, 18.0)	12.5 (−6.2, 34.9)	2.1 (−3.9, 8.4)
	$CD34^{+}$[c]	2.4 (−19.2, 29.7)	0.0 (7.04)	12.5 (−8.1, 37.7)	20.8 (−15.5, 72.8)	9.5 (−2.6, 23.2)
	$CD34^{+}CD38^{-}$[c]	−1.1 (−23.6, 28.1)	0.0 (3.11)	19.1 (−4.3, 48.1)	**47.9 (0.4, 117.8)**	**15.5 (1.6, 31.3)**
	$CD34^{+}$c-*kit*$^{+}$	13.7 (−14.0, 50.4)	0.0 (5.80)	18.9 (−5.5, 49.6)	34.2 (−10.4, 101.1)	10.0 (−4.0, 26.1)
	CFU-GM[e]	**37.0 (2.1, 83.8)**	0.0 (4.04)	29.0 (−0.1, 66.6)	31.4 (−16.2, 106.1)	6.1 (−8.4, 23.0)
Adjusted for core covariates and IGF-1	TNC	0.6 (−10.0, 12.5)	0.0 (15.00)	9.7 (−0.4, 20.8)	12.3 (−5.5, 33.4)	4.5 (−1.5, 11.0)
	MNC	4.0 (−8.0, 17.7)	0.0 (6.91)	8.5 (−2.5, 20.6)	16.8 (−3.5, 41.4)	3.7 (−2.9, 10.8)
	$CD34^{+}$[c]	11.0 (−12.3, 40.6)	0.0 (7.04)	2.5 (−16.5, 25.8)	−0.3 (−31.0, 44.0)	0.6 (−11.4, 14.2)
	$CD34^{+}CD38^{-}$[c]	5.2 (−19.0, 36.6)	0.0 (3.11)	10.9 (−11.5, 38.8)	27.6 (−14.7, 90.9)	7.9 (−6.2, 24.0)
	$CD34^{+}$c-*kit*$^{+}$	22.8 (−6.9, 62.1)	0.0 (5.80)	7.0 (−15.3, 35.2)	10.5 (−26.9, 66.9)	0.7 (−12.9, 16.4)
	CFU-GM[e]	**45.6 (8.6, 95.1)**	0.0 (4.04)	15.5 (−11.6, 50.9)	11.5 (−29.8, 77.1)	−1.7 (−16.2, 15.3)

[a]Unadjusted geometric means. TNC, initial total nucleated cells × 10^6 per ml; MNC, initial mononuclear cells × 10^6 per ml; the unit for the stem cell populations ($CD34^{+}$, $CD34^{+}CD38^{-}$, $CD34^{+}$c-*kit*$^{+}$, and CFU-GM) was per 1000 MNC. [b]Core covariates included mother's age, race of parents (both Caucasian or not), parity, gestation duration, gender of baby (male or female), delivery time (night or day) and study site (ACBP or T-NEMC). [c]$n = 233$ with complete information on all the covariates. [d]Determined only in the T-NEMC-derived samples. [e]Data from the T-NEMC-derived samples on which this assay was conducted. Statistical significance of $P < 0.05$ are given in bold.

Table 3 Ethnic-specific, core covariate-adjusted[a] multiple linear regression analysis for the association between measurements of haematopoietic stem cell populations and birth weight

		Birth weight (g) in categories				Birth weight per 500 g
Ethnicity	**Stem cell measurements**	**<3000 % Difference (95% CI)**	**3000–3499 Reference (geometric mean[b])**	**3500–3999 % Difference (95% CI)**	**⩾4000 % Difference (95% CI)**	**% Difference (95% CI)**
Both parents Caucasian	$CD34^{+}$	−13.9 (−37.2, 18.0)	0.0 (7.73)	8.3 (−15.2, 38.3)	37.5 (−11.2, 113.1)	**16.6 (0.9, 34.7)**
	$CD34^{+}CD38^{-}$	−18.0 (−43.0, 17.9)	0.0 (3.41)	14.3 (−13.7, 51.4)	**80.5 (9.1, 198.7)**	**23.8 (4.7, 46.4)**
	$CD34^{+}$c-*kit*$^{+}$	−11.5 (−40.8, 32.3)	0.0 (6.44)	18.4 (−11.0, 57.5)	63.4 (−1.0, 169.5)	**22.9 (3.6, 45.8)**
	CFU-GM	19.3 (−25.9, 92.8)	0.0 (4.01)	36.9 (−3.4, 94.1)	40.0 (−25.3, 162.5)	19.7 (−3.3, 48.2)
Either parent non-Caucasian	$CD34^{+}$	28.4(−12.1, 87.6)	0.0 (6.47)	19.9 (−16.6, 72.3)	−0.6 (−46.6, 85.3)	0.5 (−18.1, 23.5)
	$CD34^{+}CD38^{-}$	23.2 (−16.2, 81.2)	0.0 (2.84)	28.2 (−10.8, 84.4)	13.5 (−39.1, 111.7)	6.3 (−13.8, 31.0)
	$CD34^{+}$c-*kit*$^{+}$	42.1 (−6.2, 115.5)	0.0 (5.39)	22.2 (−17.2, 80.4)	9.2 (−43.9, 112.6)	−0.4 (−20.4, 24.7)
	CFU-GM	**68.4 (14.0, 148.6)**	0.0 (4.08)	21.0 (−18.1, 78.6)	24.3 (−37.5, 147.1)	−7.3 (−26.0, 16.1)

[a]Core covariates included mother's age, parity, gestation duration, gender of baby (male or female), delivery time (night or day) and study site (ACBP or T-NEMC). [b]Unadjusted geometric means. The unit for the stem cell measurements was per 1000 MNC. Statistical significance of $P < 0.05$ are given in bold.

expansion of stem cell pools result in infants with high birth weights; the larger the stem cell pool, the greater the risk that one of these stem cells will undergo malignant transformation. In support of the first tenet of this hypothesis, we have demonstrated that the concentrations of stem and progenitor cell populations in umbilical cord blood, serving as surrogates for overall stem cell potential, correlate with cord blood plasma levels of certain mitogens, notably IGF-1 (Savarese *et al*, 2007). The hypothesis also predicts that newborns with high birth weights should have elevated stem cell populations. Our findings indicate that there is a positive association between birth weight and haematopoietic stem cell measurements in the cord blood samples among newborns with normal-to-high (⩾3000 g) birth weights. This association is strongest with $CD34^{+}CD38^{-}$ cells, a relatively primitive haematopoietic stem cell population. These data are in line with previous studies, which showed a positive relationship between cord blood $CD34^{+}$ or CFU-GM levels and birth weight (Shlebak *et al*, 1998; Ballen *et al*, 2001; Aroviita *et al*, 2004).

However, in our study, newborns in the lowest birth weight category (<3000 g) can have higher levels of stem/progenitor cell measurements than those with 3000–3499 g birth weight resulting in a J- or U-shaped relation between stem cell levels and birth weight. This finding is intriguing, as J- or U-shaped relationships have often been observed in childhood cancers (Schüz and Forman, 2007) and neurological (Schüz *et al*, 2001; Von Behren and Reynolds, 2003), prostate (Eriksson *et al*, 2007), colorectal (Nilsen *et al*, 2005) and early-onset breast cancers (Sanderson *et al*, 1996; Innes *et al*, 2000; Mellemkjaer *et al*, 2003).

The possible elevated levels of stem cells at low birth weight warrant further investigations. Using birth weight to define small-for-gestation for full-term healthy infants, a majority (>86%) of such infants undergo accelerated growth during the first 6–12 months after birth and attain normal height later in life (Karlberg and Albertsson-Wikland, 1995). Premature infants (i.e., those born before the 33rd gestational week, generally with low birth weights) have been shown to have elevated foetal and cord blood $CD34^{+}$ and $CD34^{+}CD38^{-}$ levels relative to full-term infants (most with a normal birth weight) (Shields and Andrews, 1998; Wyrsch *et al*, 1999). Relevant to this phenomenon, it has been reported that premature female newborns have an increased risk for breast cancer later in life (Ekbom *et al*, 2000). It is not known whether premature or small newborns harbour elevated stem/progenitor

Epidemiology

cell populations, because these populations have not had enough developmental time to undergo a normal course of differentiation, or if such infants build up a stem cell reserve for growth compensation during the postnatal period. In any case, our J-shaped relation observed between birth weight and stem cell measurements requires further confirmation as ethnicity-specific results showed a linear relation in the considerably larger group of Caucasian cord blood samples.

Finally, we examined cord blood plasma IGF-1 levels in relation to stem cell levels and birth weight. Insulin-like growth factor-1 had a positive association with stem cell measurements in our samples (Savarese *et al*, 2007) and was strongly associated with birth weight (geometric means were 41.18, 57.86, 78.45 and 102.77 ng ml^{-1}, respectively, for the four categories of birth weight in Table 2 and in the literature (Bennett *et al*, 1983; Fant *et al*, 1993; Reece *et al*, 1994)). The growth hormone/IGF-1 axis has been suggested to serve as a master developmental regulator, coordinating stem cells in multiple organs (Ginestier and Wicha, 2007). When we controlled for cord plasma IGF-1 levels, the associations were considerably weakened. This would be expected if IGF-1 regulates stem cell potential and this is, in turn, a determinant of birth weight. To determine whether birth weight is an accurate reflection of stem cell potential, the focus should be on the results without adjusting for IGF-1.

We conclude that there is a J- or U-shaped positive association between birth weight and stem cell measurements, linear among term newborns with normal-to-high birth weight and stronger with the more primitive stem cell sub-population. The nonlinear relationship, however, suggests that birth weight is not an unequivocal indicator of stem cell potential in the context of a prenatal origin of cancer risk. Quantitation of stem cell pools might prove to be a more accurate predictor of cancer risk than birth weight *per se*, in accordance with the stem cell burden hypothesis.

ACKNOWLEDGEMENTS

We are grateful for the effort of the attending physicians and residents of the Department of Obstetrics and Gynecology and Family Practice and the staff on the Labor and Delivery ward at the Tufts-New England Medical Center. This study was support by a Grant (R01CA90902) from the US National Institutes of Health. PL, HOA, DT and CCH were supported in part by the W81XWH-05-1-0314 Innovator Award, USA. Department of Defense Breast Cancer Research Program, Office of the Congressionally Directed Medical Research Programs.

REFERENCES

Abboud M, Xu F, LaVia M, Laver J (1992) Study of early hematopoietic precursors in human cord blood. *Exp Hematol* **20:** 1043 – 1047

Adami H-O, Persson I, Ekbom A, Wolk A, Ponten J, Trichopoulos D (1995) The aetiology and pathogenesis of human breast cancer. *Mutat Res* **333:** 29 – 35

Ahlgren M, Wohlfahrt J, Olsen LW, Sørensen TIA, Melbye M (2007) Birth weight and risk of cancer. *Cancer* **110:** 412 – 419

Aroviita P, Teramo K, Hiilesmaa V, Westman P, Kekomaki R (2004) Birthweight of full-term infants is associated with cord blood CD34+ cell concentration. *Acta Paediatr* **93:** 1323 – 1329

Ballen KK, Wilson M, Wuu J, Ceredona AM, Hsieh C, Stewart FM, Popovsky MA, Quesenberry PJ (2001) Bigger is better: maternal and neonatal predictors of hematopoietic potential of umbilical cord blood units. *Bone Marrow Transplant* **27:** 7 – 14

Bennett A, Wilson DM, Liu F, Nagashima R, Rosenfeld RG, Hintz RL (1983) Levels of insulin-like growth factors I and II in human cord blood. *J Clin Endocrinol Metab* **57:** 609 – 612

Ekbom A, Erlandsson G, Hsieh CC, Trichopoulos D, Adami H-O, Cnattingius S (2000) Risk of breast cancer in prematurely born women. *J Natl Cancer Inst* **92:** 840 – 841

Eriksson M, Wedel H, Wallander M-A, Krakau I, Hugosson J, Carlsson S, Svärdsudd K (2007) The impact of birth weight on prostate cancer incidence and mortality in a population-based study of men born in 1913 and followed up from 50 to 85 years of age. *The Prostate* **67:** 1247 – 1254

Fant M, Salafia C, Baxter RC, Schwander J, Vogel C, Pezzullo J, Moya F (1993) Circulating levels of IGFs and IGF binding proteins in human cord serum: relationships to intrauterine growth. *Regul Pept* **48:** 29 – 39

Ginestier C, Wicha MS (2007) Mammary stem cell number as a determinate of breast cancer risk. *Breast Cancer Res* **9:** 109

Hoffbrand AV, Pettit JE, Moss P (2001) *Essential Haematology* 4th edn, pp 1 – 11. Malden, MA: Blackwell Science

Innes K, Byers T, Schymura M (2000) Birth characteristics and subsequent risk for breast cancer in very young women. *Am J Epidemiol* **152:** 1121 – 1128

Karlberg J, Albertsson-Wikland K (1995) Growth in full-term small-for-gestational-age infants: from birth to final height. *Pediatr Res* **38:** 733 – 739

Mayani H, Lansdorp PM (1998) Biology of human umbilical cord blood-derived hematopoietic stem/progenitor cells. *Stem Cells* **16:** 153 – 165

Mellemkjaer L, Olsen ML, Sorensen HT, Thulstrup AAM, Olsen J, Olson JH (2003) Birth weight and risk of early-onset breast cancer (Denmark). *Cancer Causes Control* **14:** 61 – 64

Michels KB, Xue F (2006) Role of birth weight in the etiology of breast cancer. *Int J Cancer* **119:** 2007 – 2025

Nilsen TIL, Romundstad PR, Troisi R, Potischman N, Vatten LJ (2005) Birth size and colorectal cancer risk: a prospective population based study. *Gut* **54:** 1728 – 1732

Reece EA, Wiznitzer A, Le E, Homko CJ, Behrman H, Spencer EM (1994) The relation between human fetal growth and fetal blood levels of insulin-like growth factors I and II, their binding proteins and receptors. *Obstet Gynecol* **84:** 88 – 95

Sanderson M, Williams MA, Malone KE, Stanford JL, Emanuel I, White E, Daling JR (1996) Perinatal factors and risk of breast cancer. *Epidemiology* **7:** 34 – 37

Savarese TM, Strohsnitter WC, Low HP, Liu Q, Baik I, Okulicz W, Chelmow DP, Lagiou P, Quesenberry PJ, Noller KL, Hsieh CC (2007) Correlation of umbilical cord blood hormones and growth factors with stem cell potential: implications for the prenatal origin of breast cancer hypothesis. *Breast Cancer Res* **9:** R29

Schüz J, Forman MR (2007) Birthweight by gestational age and childhood cancer. *Cancer Causes Control* **18:** 655 – 663

Schüz J, Kaletsch U, Meinert R, Kaatsch P, Spix C, Michaelis J (2001) Risk factors for neuroblastoma at different stages of disease Results from a population-based case – control study in Germany. *J Clin Epidemiol* **54:** 702 – 709

Sharkey AM, Jokhi PP, King A, Loke YW, Brown KD, Smith SK (1994) Expression of c-kit and kit ligand at the human maternofetal interface. *Cytokine* **6:** 195 – 205

Shields LE, Andrews RG (1998) Gestational age changes in circulating CD34+ hematopoietic stem/progenitor cells in fetal cord blood. *Am J Obstet Gynecol* **178:** 931 – 937

Shlebak AA, Roberts IAG, Stevens TA, Syzdlo RM, Goldman JM, Gordon MY (1998) The impact of antenatal and perinatal variables on cord blood haemopoietic stem/progenitor cell yield available for transplantation. *Br J Haematol* **103:** 1167 – 1171

Trichopoulos D (1990) Does breast cancer originate *in utero*? *Lancet* **335:** 939 – 940

Trichopoulos D, Lagiou P, Adami HO (2005) Towards an integrated model for breast cancer etiology: the crucial role of the number of mammary tissue-specific stem cells. *Breast Cancer Res* **7:** 13 – 17

Von Behren J, Reynolds P (2003) Birth characteristics and brain cancers in young children. *Int J Epidemiol* **32:** 248 – 256

Wyrsch A, dale Carbonare V, Jansen W, Chklovskaia E, Nissen C, Surbek D, Holzgreve W, Tichelli A, Wodnar-Filipowicz A (1999) Umbilical cord blood from preterm human fetuses is rich in committed and primitive hematopoietic progenitors with high proliferative and self-renewal capacity. *Exp Hematol* **27:** 1338 – 1345

Xiao M, Dooley DC (2000) Cellular and molecular aspects of human $CD34^{+}CD38^{-}$ precursors: analysis of a primitive hematopoietic population. *Leuk Lymphoma* **38:** 489 – 497

Epidemiology

Int. J. Cancer: **122,** 481–485 (2008)

MINI REVIEW

Early life events and conditions and breast cancer risk: From epidemiology to etiology

Dimitrios Trichopoulos[1]*, Hans-Olov Adami[1,2], Anders Ekbom[3], Chung-Cheng Hsieh[1,4] and Pagona Lagiou[1,5]

[1]*Department of Epidemiology, Harvard School of Public Health, Boston, MA*
[2]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
[3]*Department of Medicine, Unit of Clinical Epidemiology, Karolinska Institutet/Karolinska University Hospital, Stockholm, Sweden*
[4]*Department of Cancer Biology and Cancer Center, University of Massachusetts Medical School, Worcester, MA*
[5]*Department of Hygiene and Epidemiology, School of Medicine, University of Athens, Greece*

Risk factors for breast cancer—documented by intensive epidemiological investigations and viewed in the context of general principles of carcinogenesis—can be integrated to an etiologic model comprising 3 principal components: the likelihood of breast cancer occurrence depends on the number of mammary tissue-specific stem cells, which is determined in early life; all growth-enhancing mammotropic hormones affect the rate of expansion of initiated clones; and while a pregnancy stimulates the replication of already initiated cells, it conveys long-term protection through differentiation of mammary tissue-specific stem cells. This perspective accommodates much of what is known about the epidemiology and natural history of breast cancer and highlights the role of early life in the origin of this cancer.

Key words: breast cancer; perinatal; birth weight; stem cells; hormones; pregnancy

Breast cancer epidemiology

The incidence of breast cancer has apparently increased throughout the world during the last century, even before the widespread application of mammographic screening programs and mortality from the disease in developed countries generally exceeds that from other cancer sites.[1] Breast cancer epidemiology has been intensively studied, perhaps more than that of any other cancer.[2–6] Table I summarizes what are generally considered as established epidemiological characteristics of breast cancer and provides an indication of the strength of the respective associations in terms of the relative risk per natural contrasts or usual increments.

Breast cancer is mostly, though not exclusively, a disease of women. The incidence of the disease increases with age, with an inflection around menopause, which is not evident for other forms of cancer. It is generally more common among urban rather than rural residents as well as among women of higher socioeconomic status. In comparison to Asian women in China or Japan, Caucasian women in the western world have a considerably higher breast cancer risk.[1]

With respect to reproductive history, an earlier age at menarche and a later age at menopause are associated with increased risk whereas, for a given age at menopause, induced menopause conveys more protection than the naturally occurring one.[6–8] The role of pregnancies is complex. Irrespective of the woman's age, a pregnancy imparts a short-term increase of breast cancer risk[9] followed by a substantial long-term reduction of this risk, as was first documented with respect to the first pregnancy some 40 years ago in a classical international epidemiological study.[10] Hence, the earlier the age at first full-term pregnancy, the more prolonged is the subsequent long-term protection. After the age of about 35 years, a first pregnancy actually increases breast cancer risk, apparently because the short-term risk increase exceeds the subsequent risk reduction. Additional full-term pregnancies have similar but much weaker effects,[11] while spontaneous or induced abortions do not affect breast cancer risk.[12] Prolonged lactation conveys at most modest protection, which appears to be restricted to premenopausal women.[13,14] Current or recent use of oral contraceptives slightly increase the risk for breast cancer,[15] whereas long-term use of replacement estrogens with progestins may substantially increase breast cancer risk.[16–18]

High birth weight has been associated with increased breast cancer risk in the offspring.[19] Having been breastfed as an infant has been investigated for its role in breast cancer under the assumption that it could be responsible for the transmission of an infectious agent, but the results did not support an association.[20] Early life growth[21] and factors that may increase it[22] have also been positively associated with breast cancer risk, as has height[23,24] and post- (but not pre-) menopausal obesity[8,25–27] later in life.

A high-density mammogram (75% or more of the total breast area with dense mammographic appearance) has been associated with a more than 4-fold risk in comparison to a low-density mammogram (10% or less or total breast area with dense mammographic appearance).[28] Atypical hyperplasia of the mammary gland has been documented as an important breast cancer risk factor.[29,30]

Family history among first degree relatives is associated with increased breast cancer risk.[31] BRCA1 and BRCA2, as well as some highly penetrant mutations, explain a large part of familial breast cancers, but account for a small proportion of all breast cancers.[32] Many studies have examined low penetrance susceptibility polymorphisms in candidate genes, but the associations reported in some studies could not been replicated in subsequent investigations. This is an evolving field, in which large whole genome association investigations are providing new insight.[33] Breast cancer in the contralateral breast is an established risk factor for developing the disease in the other breast, but the underlying pathogenetic mechanisms are not clear.[34]

High levels of physical activity[35] and high intake of vegetables, perhaps fruits[36] and olive oil[37] have been reported to be associated with reduced breast cancer risk, possibly by reducing endogenous estrogen levels.[38,39] Nevertheless, the evidence is inconclusive and suggests, at most, weak effects. Recent evidence points to total and particularly saturated fat as been weakly, but significantly, positively associated with breast cancer risk.[40] Most studies indicate that consumption of alcoholic beverages may slightly

---

Based on the Third Richard Doll lecture given at the International Agency for Research on Cancer by DT on May 9, 2007 and relying on the collective work of all authors.

Grant sponsors: W81XWH-05-1-0314 Innovator Award, US. Department of Defence Breast Cancer Research Program, Office of the Congressionally Directed Medical Research Programs.

***Correspondence to:** Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. Fax: +1-617-566-7805. E-mail: dtrichop@hsph.harvard.edu

Received 24 October 2007; Accepted after revision 25 October 2007
DOI 10.1002/ijc.23303
Published online 16 November 2007 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – FACTORS EVALUATED IN RELATION TO BREAST CANCER RISK

Risk factor	Category/change	Strength
Gender	Women vs. men	++++
Age	Increase	++++
Ethnic group	Caucasian vs. Asian	+++
Family history	Yes vs. no	+++
Specific genes	Yes vs. no	++++
Cancer in other breast	Yes vs. no	+++
Height	Increase	++
Postmenopausal obesity	Increase	++
Birth weight	Increase	+
Having been breastfed	No vs. yes	0
Growth in early life	Increase	+
Atypical hyperplasia	Present vs. absent	+++
Mammographic density (mammary gland mass)	High vs. low density (increasing mass)	+++
Age at menarche	Earlier	++
Age at menopause	Later	++
Type of menopause	Natural vs. artificial	++
Age at 1st full term pregnancy	Later	+++
Age at other pregnancies	Later	+
Parity overall	Lower	++
Pregnancy timing	Proximal vs. distant	+
Lactation	No vs. yes	+
Abortion	No vs. yes	0
Oral contraceptive use (recent)	Increase	+
Hormone replacement	Increase	++
Plant foods and olive oil	Reduced intake	+
Saturated fat	Increased intake	+
Physical activity	Reduced	+
Ethanol intake	Increase	+
Ionizing radiation	Increased	+
Magnetic fields	Increased	0
Organochlorines	Increased	0

Association: ++++ very strong, +++ strong, ++ modest, + weak, 0 null.

increase breast cancer risk, possibly by increasing estrogen levels.[41] There is no conclusive evidence for an association between tobacco smoking and the disease.[42,43]

Ionizing radiation is an established cause of cancer of the breast as well as of several other cancers, but it is of limited quantitative importance. Exposure to organochlorines[44] or electromagnetic fields[45] has not been shown to be related to breast cancer.

It is generally believed that the association between endogenous hormones and breast cancer risk should be studied in prospective, rather than retrospective, investigations, under the undocumented assumption that disease status, even prior to treatment, may affect hormone levels. Among postmenopausal women, most hormones examined—with the notable exception of adiponectin that has been mostly evaluated through case-control designs[46,47]—have been positively associated with breast cancer risk.[6,48–50] The list includes total and free estradiol, estrone and estrone sulphate, androstenedione, dehydroepiandrosterone and dehydroepiandrosterone sulphate, testosterone and prolactin. Among premenopausal women, case-control studies and a few cohort investigations provide some support for a positive association between estrogens and breast cancer risk, but they also indicate that high levels of androgens could increase this risk.[6,51] In both prospective and retrospective studies among premenopausal women, significant positive associations have been reported between blood insulin like growth factor 1 (IGF-1) and breast cancer risk.[52]

The early life etiological model

The etiological model we have proposed for breast cancer accommodates most of what we know about the epidemiology of the disease. The model emphasizes early life events and conditions as determinants of breast cancer risk and summarizes the distinct epidemiological characteristics of the disease on the basis of 3 major components[5,22,53–58]:

- The likelihood of breast cancer occurrence depends on the number of mammary tissue-specific stem cells, which is determined early in life, including the intrauterine life,
- in early and later life, growth-enhancing mammotropic hormones affect the replication rate of mammary tissue specific stem cells, the likelihood of retention of cells with spontaneous somatic mutations as well as the rate of expansion of initiated clones, and
- while a pregnancy stimulates the replication of already initiated cells, it conveys long-term protection through differentiation of a large fraction of the mammary tissue-specific stem cells.

It should be noted that several scientists[23,59,60] have postulated, explicitly or implicitly, that early life influences may play a role in breast cancer etiology and there have even been early studies exploring birth weight as a breast cancer risk factor.[61] Moreover, the issue of pregnancy induced terminal differentiation of mammary gland has been championed by Russo and Russo.[62]

Categorization of breast cancer risk factors according to the 3 components of the early life etiological model

An etiological model should accommodate the epidemiological profile of the disease it aims to explain. In this context, we have categorized the established breast cancer risk factors according to the 3 components of the early life etiological model, taking also into account that certain breast cancer epidemiologic characteristics reflecting general principles of carcinogenesis relevant to many cancer sites (Table II). The empirical evidence in support of the categorization has been presented in detail in earlier publications[2,5] and is summarized below.

First component

Mammary gland mass, *as distinct from breast size*, is usually assessed through mammographic density and is an important breast cancer risk factor.[28] Mammary gland mass, which is likely to reflect the pool of mammary cells and be correlated with the number of mammary stem cells,[63,64] can also accommodate several breast cancer risk factors, including the higher incidence of the disease among Caucasian compared to Asian women and women of higher rather than lower socioeconomic class as well as the preponderance of breast cancer in the slightly larger left, rather than right, breast.[65] The positive associations of breast cancer risk with birth weight, growth in early life and adult height could also be explained in terms of mammary gland mass. Finally, at the extreme, the strikingly higher breast cancer risk among women than among men, even in later life when estrogen production is not substantially different between the 2 genders, is best explained by the correspondingly higher mammary gland mass among women than among men.

Second component

Most investigators agree that oestrogens in general, or specific categories of oestrogens, or prolactin, or other hormones, including IGF, are important in the etiology of breast cancer. Our view is that all growth enhancing and mammotropic hormones are involved in one or more stages in the long process leading to clinical breast cancer. An important issue that has not been sufficiently explored in empirical research is the way these hormones interact in the causation of the disease. A small study presented evidence that mammotropic hormones may act as permissive factors for breast cancer occurrence and that values of any one of these above a certain level may suffice for sustaining growth of a developing tumor[66]—the finding is intriguing but requires confirmation in larger datasets. The second component of the etiologic model accommodates our knowledge about the role of reproductive factors in the etiology of the disease

TABLE II – GROUPING OF BREAST CANCER RISK FACTORS ACCORDING TO THE GENERAL PRINCIPLES OF CARCINOGENESIS AND THE POSTULATED PATHOGENIC PROCESS

General principles of carcinogenesis	Number of mammary tissue specific stem cells	Growth enhancing mammotropic hormones	Terminal differentiation
Age	Mammographic density (gland mass)	Gender	Age at 1st full term pregnancy
Ionizing radiation	Atypical hyperplasia	Age incidence pattern	Age at other pregnancies
Family history	Gender	Age at menarche	Parity overall
Specific genes	Birth weight	Age at menopause	Lactation
	Growth in early life	Type of menopause	
	Height	Oral contraceptives	
	Ethnic group	Hormone replacement	
		Pregnancy timing	
		Postmenopausal obesity	
		Ethanol intake	
		Physical activity	
		Adult life diet	

as well as that of alcohol drinking (which tends to increase oestrogen levels), physical activity and adult life diet.[2,5]

Third component

Terminal differentiation of the mammary gland takes place mostly after the occurrence of the first full-term pregnancy, and to a lesser extent, after the occurrence of subsequent pregnancies and lactation.[67] The later the age at first full term pregnancy, the higher the number of already initiated cells and the more limited the protection conveyed by pregnancy. Beyond the age of 35 or so, the transient increase of breast cancer risk that accompanies a pregnancy (due to the effect on already initiated clones of the many-fold increases of mammotropic and growth enhancing hormones) overshadows the protection conveyed by the terminal differentiation of immature mammary cells. The 3rd component of the etiologic model also accommodates what was largely thought to be an enigma, namely why breast cancer risk is higher among parous than among nulliparous women of premenopausal age.

The ecological challenges

One of the most challenging characteristics in breast cancer epidemiology is the sharp ecological contrast in breast cancer incidence between women in western Europe and North America and women in China and Japan,[1] which fades in Asian women migrating to the west after 2 or more generations. Neither reproductive nor dietary factors in adult life can explain the 4-fold difference in incidence observed in these populations,[8,68] nor can they explain the subsequent incidence assimilation. On the contrary, diet in early life could provide an explanation for the ecological contrast in the context of the early life etiological model: reduced energy intake in early life is associated with smaller body size in adult life and smaller body size constrains birth weight and subsequent development of offspring. Increased energy intake, on the other hand, facilitates growth and removes constrains on birth weight and eventual body size. This cycle tends to repeat over consecutive generations of Asian migrants in western countries and is associated with a gradual increase in body size and breast cancer incidence among them.[22,57]

The early life etiological model is not refuted by the fact that populations at low risk for breast cancer have higher levels of most pregnancy—or possibly adult life—hormones.[69] It is plausible that in striking ecological contrasts (*e.g.* between native Chinese and Caucasian populations), pregnancy growth hormones tend to increase in order to compensate for physically constrained fetal growth[58,70] and the perinataly programmed higher levels of these hormones could track through adult life.

Avenues of future research

Future research assessing the early life aspects of the etiology of breast cancer could follow many directions and some of them are outlined here. Hsieh and coworkers[65,71,72] are evaluating how pregnancy mammotropic and growth hormones affect cord blood stem cell populations. In their recent work, they reported that cord blood plasma levels of IGF-1 were strongly correlated with all the hematopoietic stem and progenitor concentrations examined, whereas estriol and insulin-like growth factor binding protein-3 levels were positively and significantly correlated with some of these cell populations. Hilakivi-Clarke and her coworkers have used rodent models to explore the ways through which diet and otherwise induced epigenetic changes in target genes might lead to strategies to prevent breast cancer.[73,74] Critically important results may also emerge from a unique follow-up of women born to mothers who have taken diethylstilbestrol (DES) during their pregnancies. Two recent publications indicated that in utero DES exposure may substantially increase breast cancer risk in the offspring.[75,76] Moreover, it would be important to firmly document what has already been reported in previous publications, that is, that perinatal characteristics predictive of high breast cancer risk in adult life are also predictive of high breast cancer risk mammographic patterns.[77,78]

Conclusion

The early life etiologic model we have outlined accommodates the existing epidemiological evidence. Its 3 components refer to stages of a single biological process that points to the number of mammary tissue-specific stem cells as a core determinant of breast cancer risk. The first component focuses on the perinatal period, when stem cells are generated. The second component concentrates on preinitiation and postinitiation growth factors that modulate the number of mammary stem cells at risk and the growth of the initiated clones. The third postulate explains how cells at risk are removed through terminal differentiation.

References

1. Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2002 cancer incidence, mortality and prevalence worldwide. IARC CancerBase No. 5, version 2.0. Lyon: IARC, 2004.
2. Adami H-O, Signorello LB, Trichopoulos D. Towards an understanding of breast cancer etiology. Seminars in Cancer Biology 1998:8: 255–262.
3. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet Oncol 2001;2:133–40.
4. Richie RC, Swanson JO. Breast cancer: a review of the literature. J Insur Med 2003;35:85–101.
5. Trichopoulos D, Lagiou P, Adami HO. Towards an integrated model for breast cancer etiology: the crucial role of the number of mammary tissue-specific stem cells. Breast Cancer Res 2005;7:13–17.
6. Hankinson S, Tamimi R, Hunter D. Breast cancer. In: Adami HO, Hunter D, Trichopoulos D, eds. Textbook of cancer epidemiology. New York: Oxford University Press, in press.

7. Trichopoulos D, MacMahon B, Cole P. Menopause and breast cancer risk. J Natl Cancer Inst 1972;48:605–13.
8. Hsieh C-c, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. Int J Cancer 1990;46:796–800.
9. Lambe M, Hsieh C-c, Trichopoulos D, Ekbom A, Pavia M, Adami H-O. Transient increase in the risk of breast cancer after giving birth. N Engl J Med 1994;331:5–9.
10. MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, Salber EJ, Valaoras VG, Yuasa S. Age at first birth and breast cancer risk. Bull World Health Organ 1970;43:209–21.
11. Trichopoulos D, Hsieh C, MacMahon B, Lin T, Lowe C, Mirra A, Ravnihar B, Salber E, Valaoras V, Yuasa S. Age at any birth and breast cancer risk. Int J Cancer 1983;31:701–4.
12. Melbye M, Wohlfahrt J, Olsen JH, Frisch M, Westergaard T, Helweg-Larsen K, Andersen PK. Induced abortion and the risk of breast cancer. N Engl J Med 1997;336:81–5.
13. MacMahon B, Lin TM, Lowe CR, Mirra AP, Ravnihar B, Salber EJ, Trichopoulos D, Valaoras VG, Yuasa S. Lactation and cancer of the breast. A summary of an international study. Bull World Health Organ 1970;42:185–94.
14. Newcomb PA, Storer BE, Longnecker MP, Mittendorf R, Greenberg ER, Clapp RW, Burke KP, Willett WC, MacMahon B. Lactation and a reduced risk of premenopausal breast cancer. N Engl J Med 1994;330:81–7.
15. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53297 women with breast cancer and 100239 women without breast cancer from 54 epidemiological studies. Lancet 1996;347:1713–27.
16. Bergkvist L, Adami HO, Persson I, Hoover R, Schairer C. The risk of breast cancer after estrogen and estrogen-progestin replacement. N Engl J Med 1989;321:293–7.
17. Beral V and Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet 2003;362:419–27.
18. Chlebowski RT, Hendrix SL, Langer RD, Stefanick ML, Gass M, Lane D, Rodabough RJ, Gilligan MA, Cyr MG, Thomson CA, Khandekar J, Petrovitch H, McTiernan A, WHI Investigators. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. JAMA 2003;289:3243–53.
19. Michels KB, Xue F. Role of birthweight in the etiology of breast cancer. Int J Cancer 2006;119:2007–25.
20. Titus-Ernstoff L, Egan KM, Newcomb PA, Baron JA, Stampfer M, Greenberg ER, Cole BF, Ding J, Willett W, Trichopoulos D. Exposure to breast milk in infancy and adult breast cancer risk. J Natl Cancer Inst 1998;90:921–4.
21. Ahlgren M, Melbye M, Wohlfahrt J, Sørensen TI. Growth patterns and the risk of breast cancer in women. N Engl J Med 2004;351:1619–26.
22. Lagiou P, Adami HO, Trichopoulos D. Early life diet and the risk for adult breast cancer. Nutr Cancer 2007;56:158–61.
23. Valaoras V, MacMahon B, Trichopoulos D, Polychronopoulou A. Lactation and reproductive histories of breast cancer patients in Greater Athens, 1965–1967. Int J Cancer 1969;4:350–63.
24. Tretli S. Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570,000 women in Norway. Int J Cancer 1989;44:23–30.
25. Ursin G, Longnecker MP, Haile RW, Greenland S. A meta-analysis of body mass index and risk of premenopausal breast cancer. Epidemiology 1995;6:137–41.
26. Hunter DJ, Willett WC. Diet, body size, and breast cancer. Epidemiol Rev 1993;15:110–32.
27. Magnusson C, Baron J, Persson I, Wolk A, Bergstrom R, Trichopoulos D, Adami HO. Body size in different periods of life and breast cancer risk in post-menopausal women. Int J Cancer 1998;76:29–34.
28. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S, Yaffe MJ. Mammographic density and the risk and detection of breast cancer. New Engl J Med 2007;356:227–36.
29. Dupont WD, Parl FF, Hartmann WH, Brinton LA, Winfield AC, Worrell JA, Schuyler PA, Plummer WD. Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer 1993;71:1258–65.
30. Singletary SE. Rating the risk factors for breast cancer. Ann Surg 2003;237:474–82.
31. Colditz GA, Willett WC, Hunter DJ, Stampfer MJ, Manson JE, Hennekens CH, Rosner BA. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study. JAMA 1993;270:338–43.
32. Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. Cancer 1996;77:2318–24.
33. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
34. Hartman M, Czene K, Reilly M, Bergh J, Lagiou P, Trichopoulos D, Adami HO, Hall P. Genetic implications of bilateral breast cancer: a population based cohort study. Lancet Oncol 2005;6:377–82.
35. Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, van Leeuwen FE; TFPAC. Physical activity and breast cancer: a systematic review. Epidemiology 2007;18:137–57.
36. World Cancer Research Fund and American Institute for Cancer Prevention. Food, Nutrition and the Prevention of Cancer: a global perspective. Washington DC: World Cancer Research Fund and American Institute for Cancer Prevention, 1997;252.
37. Trichopoulou A. Olive oil and breast cancer. Cancer Causes Control 1995;6:475–6.
38. Dorgan JF, Reichman ME, Judd JT, Brown C, Longcope C, Schatzkin A, Forman M, Campbell WS, Franz C, Kahle L, Taylor PR. Relation of energy, fat, and fiber intakes to plasma concentrations of estrogens and androgens in premenopausal women. Am J Clin Nutr 1996;64:25–31.
39. Carruba G, Granata OM, Pala V, Campisi I, Agostara B, Cusimano R, Ravazzolo B, Traina A. A traditional Mediterranean diet decreases endogenous estrogens in healthy postmenopausal women. Nutr Cancer 2006;56:253–9.
40. Thiebaut AC, Kipnis V, Chang SC, Subar AF, Thompson FE, Rosenberg PS, Hollenbeck AR, Leitzmann M, Schatzkin A. Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort. J Natl Cancer Inst 2007;99:451–62.
41. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA 2001;286:2143–51.
42. Trichopoulos D, Lagiou A. Tobacco smoking and cancer of the breast. In: Boyle P, Gray N, Henningfield J, Seffrin J, Zatonski W, eds. Tobacco and public health: science and policy. Oxford: Oxford University Press, 2004;503–510.
43. Gram IT, Braaten T, Terry PD, Sasco AJ, Adami HO, Lund E, Weiderpass E. Breast cancer risk among women who start smoking as teenagers. Cancer Epidemiol Biomarkers Prev 2005;14:61–6.
44. Laden F, Hankinson SE, Wolff MS, Colditz GA, Willett WC, Speizer FE, Hunter DJ. Plasma organochlorine levels and the risk of breast cancer: an extended follow-up in the Nurses' Health Study. Int J Cancer 2001;91:568–74.
45. Laden F, Neas LM, Tolbert PE, Holmes MD, Hankinson SE, Spiegelman D, Speizer FE, Hunter DJ. Electric blanket use and breast cancer in the Nurses' Health Study. Am J Epidemiol 2000;152:41–9.
46. Miyoshi Y, Funahashi T, Kihara S, Taguchi T, Tamaki Y, Matsuzawa Y, Noguchi S. Association of serum adiponectin levels with breast cancer risk. Clin Cancer Res 2003;9:5699–704.
47. Tworoger SS, Eliassen AH, Kelesidis T, Colditz GA, Willett WC, Mantzoros CS, Hankinson SE. Plasma adiponectin concentrations and risk of incident breast cancer. J Clin Endocrinol Metab 2007;92:1510–16.
48. Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606–16.
49. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PH, Biessy C, Dossus L, Lukanova A, Bingham S, Khaw KT, Allen NE, Bueno-de-Mesquita HB, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. Endocr Relat Cancer 2005;12:1071–82.
50. Adami HO, Johansson ED, Vegelius J, Victor A. Serum concentrations of estrone, androstenedione, testosterone and sex-hormone-binding globulin in postmenopausal women with breast cancer and in age-matched controls. Ups J Med Sci 1979;84:259–74.
51. Kaaks R, Berrino F, Key T, Rinaldi S, Dossus L, Biessy C, Secreto G, Amiano P, Bingham S, Boeing H, Bueno de Mesquita HB, Chang-Claude J, et al. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). J Natl Cancer Inst 2005;97:755–65.
52. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I. IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet 2004;363:1346–53.
53. Trichopoulos D. Hypothesis: does breast cancer originate in utero? Lancet 1990;335:939–40.
54. Ekbom A, Trichopoulos D, Adami H-O, Hsieh C-c, Lan S-j. Evidence of prenatal influences on breast cancer risk. Lancet 1992;340:1015–18.

55. Adami H-O, Persson I, Ekbom A, Wolk A, Pontén J, Trichopoulos D. The aetiology and pathogenesis of human breast cancer. Mutat Res 1995;333:29–35.
56. Trichopoulos D, Lipworth L. Editorial: is cancer causation simpler than we thought, but more intractable? Epidemiology 1995;6: 347–9.
57. Lagiou P, Hsieh CC, Trichopoulos D, Xu B, Wuu J, Mucci L, Tamimi R, Adami HO, Cnattingius S. Birthweight differences between USA and China and their relevance to breast cancer aetiology. Int J Epidemiol 2003;32:193–8.
58. Trichopoulos D. Intrauterine environment, mammary gland mass and breast cancer risk. Breast Cancer Res 2003;5:42–4.
59. Cole P, MacMahon B. Oestrogen fractions during early reproductive life in the aetiology of breast cancer. Lancet 1969;1:604–6.
60. DeWaard F, Trichopoulos D. A unifying concept of the aetiology of breast cancer. Int J Cancer 1988;41:666–9.
61. Le Marchand L, Kolonel LN, Myers BC, Mi MP. Birth characteristics of premenopausal women with breast cancer. Br J Cancer 1988;57: 437–9.
62. Russo J, Russo IH. Development of the human mammary gland. In: Neville MC, Daniel CW, eds. The mammary gland. New York: Plenum Press, 1987.67–93.
63. Trichopoulos D, Lipman R. Mammary gland mass and breast cancer risk. Epidemiology 1992;3:523–6.
64. Baik I, Becker PS, DeVito WJ, Lagiou P, Ballen K, Quesenberry PJ, Hsieh CC. Stem cells and prenatal origin of breast cancer. Cancer Causes Control 2004;15:517–30.
65. Ekbom A, Adami H-O, Trichopoulos D, Lambe M, Hsieh C-c, Pontén J. Epidemiologic correlates of breast cancer laterality (Sweden). Cancer Causes Control 1994;5:510–16.
66. Trichopoulos D, Bamia C, Lagiou P, Trichopoulou A. Are mammotropic hormones mainly permissive for the development of breast cancer? Int J Cancer 2006;118:2863–5.
67. Russo J, Romero AI, Russo IH. Architectural pattern of the normal and cancerous breast under the influence of parity. Cancer Epidemiol Biomark Prev 1994;3:219–24.
68. Willett W. Nutritional epidemiology, 2nd edn. New York: Oxford University Press, 1998.377–413.
69. Lipworth L, Hsieh C-c, Wide L, Ekbom A, Yu S-z, Yu G-p, Xu B, Hellerstein S, Carlstrom K, Trichopoulos D, Adami H-O. Maternal pregnancy hormone levels in an area with a high incidence (Boston, USA) and in an area with a low incidence (Shanghai, China) of breast cancer. Brit J Cancer 1999;79:7–12.
70. Lagiou P, Samoli E, Lagiou A, Hsieh CC, Adami HO, Trichopoulos D. Maternal height, pregnancy estriol and birth weight in reference to breast cancer risk in Boston and Shanghai. Int J Cancer 2005;117: 494–8.
71. Baik I, Devito WJ, Ballen K, Becker PS, Okulicz W, Liu Q, Delpapa E, Lagiou P, Sturgeon S, Trichopoulos D, Quesenberry PJ, Hsieh CC. Association of fetal hormone levels with stem cell potential: evidence for early life roots of human cancer. Cancer Res 2005;65:358–63.
72. Savarese TM, Strohsnitter WC, Low HP, Liu Q, Baik I, Okulicz W, Chelmow DP, Lagiou P, Quesenberry PJ, Noller KL, Hsieh CC Correlation of umbilical cord blood hormones and growth factors with stem cell potential: implications for the prenatal origin of breast cancer hypothesis. Breast Cancer Res 2007;9:R29.
73. De Assis S, Hilakivi-Clarke L. Timing of dietary estrogenic exposures and breast cancer risk. Ann NY Acad Sci 2006;1089:14–35.
74. Hilakivi-Clarke L, de Assis S. Fetal origins of breast cancer. Trends Endocrinol Metab 2006;17:340–8.
75. Palmer JR, Hatch EE, Rosenberg CL, Hartge P, Kaufman RH, Titus-Ernstoff L, Noller KL, Herbst AL, Rao RS, Troisi R, Colton T, Hoover RN. Risk of breast cancer in women exposed to diethylstilbestrol in utero: prelimiinary results (United States). Cancer Causes Control 2002;13:753–8.
76. Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, Kaufman R, Herbst AL, Noller KL, Hyer M, Hoover RN. Prenatal diethylstilbestrol exposure and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2006;15:1509–14.
77. Ekbom A, Thurfjell E, Hsieh CC, Trichopoulos D, Adami HO. Perinatal characteristics and adult mammographic patterns. Int J Cancer 1995;61:177–80.
78. Cerhan JR, Sellers TA, Janney CA, Pankratz VS, Brandt KR, Vachon CM. Prenatal and perinatal correlates of adult mammographic breast density. Cancer Epidemiol Biomarkers Prev 2005;14:1502–8.

Cord Serum Estrogens, Androgens, Insulin-Like Growth Factor-I, and Insulin-Like Growth Factor Binding Protein-3 in Chinese and U.S. Caucasian Neonates

Rebecca Troisi,[1,3] Pagona Lagiou,[4,5,7] Dimitrios Trichopoulos,[5,7] Biao Xu,[8] Lucy Chie,[6] Frank Z. Stanczyk,[9] Nancy Potischman,[2] Hans-Olov Adami,[5,7] Robert N. Hoover,[1] and Chung-Cheng Hsieh[5,7,10]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services; [2]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland; [3]Department of Community and Family Medicine, Dartmouth Medical School, Hanover, New Hampshire; [4]Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece; [5]Department of Epidemiology, Harvard School of Public Health; [6]Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; [7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [8]Department of Epidemiology, School of Public Health, Fudan University, Shanghai, China; [9]Reproductive Endocrine Research Laboratory, University of Southern California Keck School of Medicine, Los Angeles, California; and [10]Department of Cancer Biology, University of Massachusetts Medical Center, Worcester, Massachusetts

Abstract

Markedly lower breast cancer incidence rates in Asians than Caucasians are not explained by established adult risk factors. Migration studies suggest the importance of early-life exposures, including perhaps the *in utero* period. Concentrations of steroid hormones and insulin-like growth factors (IGF) were measured in umbilical cord sera from pregnancies in Shanghai, China (n = 121) and Boston, MA (n = 111). Pregnancy characteristics were ascertained by interview and medical records. Means and percent differences in hormone concentrations comparing Chinese with Caucasians and 95% confidence intervals were estimated from linear regression models. Cord concentrations of androstenedione (91.9%), testosterone (257%), estriol (48.6%), and IGF binding protein-3 (21.1%) were significantly higher in the Chinese than U.S. samples, and cord prolactin was lower (−14.9%). Cord estradiol and IGF-I concentrations did not differ by race/ethnicity. With adjustment for gestational length, maternal age, pre-pregnancy weight, and weight gain, androstenedione (60.5%), testosterone (185%), and IGF binding protein-3 (40.4%) remained significantly higher in the Chinese, whereas the higher estriol and lower prolactin concentrations were attenuated. In addition, estradiol levels became lower in the Chinese (−29.8%) but did not reach statistical significance. Results were generally similar when restricted to first full-term pregnancies, with reduced estradiol concentrations in the Chinese reaching statistical significance after adjustment. These data are consistent with the hypothesis that elevated prenatal androgen exposure could mediate reductions in breast cancer risk. The meaning of the change in findings for estrogens after controlling for factors related to the pregnancy is unclear with regard to explaining international breast cancer differences. (Cancer Epidemiol Biomarkers Prev 2008;17(1):224–31)

Introduction

The most pronounced variation in breast cancer rates is observed internationally. Incidence in East and Southeast Asia is nearly one-fifth of that in northern and western Europe (1), but rates gradually increase among Asian migrants to western countries (2). Breast cancer incidence rates in the first generation born in the west, however, may be substantially elevated when compared with migrants who were born in Asia but lived decades in the west (2). Furthermore, differences in incidence rates by migration status are not fully explained by established menstrual and reproductive risk factors for breast cancer (3, 4). These observations are consistent with environmental factors early in life explaining at least some of the variation in breast cancer rates across populations (5-7), and with mother's environment during pregnancy influencing the *in utero* environment.

Whereas the full range of hormones involved in breast carcinogenesis is unclear, evidence indicates the importance of estrogens, androgens, and progesterone (8), all of which rise significantly during pregnancy. Trichopoulos (9) hypothesized that exposure to lower *in utero* estrogen concentrations affords protection against subsequent breast carcinogenesis. In the nonpregnant state, circulating estrogen concentrations are generally lower in premenopausal and postmenopausal Asian women compared with Caucasian women (10, 11). Although

Received 6/12/07; revised 10/24/07; accepted 11/5/07.

Grant support: W81XWH-05-1-0314 Innovator Award, U.S. Department of Defense Breast Cancer Research Program, Office of the Congressionally Directed Medical Research Programs (P. Lagiou, D. Trichopoulos, H-O. Adami, and C-C. Hsieh).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Rebecca Troisi, Dartmouth-Hitchcock Medical Center, Room 854, 7297 Rubin Building, One Medical Center Drive, Lebanon, NH 03756. Phone: 603-653-9024. E-mail: troisir@mail.nih.gov

doi:10.1158/1055-9965.EPI-07-0536

androgens also have been generally lower in Asian women (10), a recent study showed an inverse correlation with increasing westernization in Asian migrants to the West (12). An investigation of maternal mid-pregnancy serum hormone concentrations, however, showed *higher* estradiol and estriol levels in Chinese women compared with American women (13) as well as elevations in several other compounds, including prolactin, progesterone, human growth hormone, albumin, sex hormone-binding globulin, and possibly α-fetoprotein levels (14). However, maternal hormone concentrations may or may not be representative of fetal exposure, and investigating hormone differences in the fetal circulation is warranted.

These are the first reported data on neonates born in China and in the United States to address whether cord concentrations of several estrogens and androgens, insulin-like growth factor (IGF)-I, and IGF binding protein (IGFBP)-3 differ between countries with low and high breast cancer incidence rates.

Materials and Methods

The study has been described previously (13). Pregnant women were recruited at their first prenatal visit to collaborating maternity clinics at the Beth Israel Hospital (Boston, MA) and from hospitals affiliated with the Shanghai Medical University (Shanghai, China). All U.S. women were urban residents, whereas women from Shanghai were recruited from three urban clinics and one rural clinic. The institutional review boards in Boston and Shanghai approved the study, and informed consent was obtained from all study participants.

Included were neonates of women less than 40 years old, with at most one previous still or live-born child. Only Caucasians in the United States and Chinese in Shanghai were included, and in both places, the mothers had to be proficient in the local language. Neonates born to women who had taken any hormonal medication during the index pregnancy or who had a previous diagnosis of diabetes mellitus or thyroid disease were excluded from the study, as were those neonates with a known major anomaly.

Between March 1994 and October 1995, 402 eligible women were identified at Beth Israel Hospital. Of these women, 77 (19.2%) declined to participate. An additional 9 (2.2%) women were excluded at a later date because of early spontaneous or induced pregnancy termination, 2 (0.5%) because of a twin birth, and 10 (2.5%) were lost to follow-up after the initial meeting. In Shanghai, 424 eligible women were identified between April 1994 and May 1995. Of these, 73 (17.2%) declined to participate, 2 (0.5%) were later excluded because of induced abortion, 2 (0.5%) because of a twin birth, 5 (1.2%) because of implied gestation durations of <30 or >50 weeks, and 7 (1.7%) were lost to follow-up after the initial meeting. In total, 304 and 335 pregnant women were enrolled in the study from Boston and Shanghai, respectively.

Umbilical cord blood collection started in December 1994. At delivery, the placenta was weighed, with the cord cut at the insertion site and the extra blood and clots minimized. Mixed cord blood was collected in sterile tubes without preservatives and refrigerated at 4°C for up to 24 hours until centrifugation. Samples were transported in a cooler from the rural clinic in China to a laboratory near Shanghai Medical University where they were centrifuged on the same day and the serum was aliquoted. The aliquots were stored at −20°C for about 5 to 7 days in the laboratory before being transported to Shanghai Medical University and stored at −80°C with the samples collected at the Shanghai hospitals. At the end of the study, all samples were shipped by air on dry ice to Boston where they were stored at −80°C. In total, 246 (115 from Boston and 131 from Shanghai) cord blood samples were collected.

Analytes were measured in cord serum at the Reproductive Endocrine Research Laboratory of the University of Southern California Keck School of Medicine under the direct supervision of one of us (F.Z.S.). Levels of estradiol, testosterone, and androstenedione were measured by RIA following extraction with organic solvent and purification by Celite column partition chromatography (15-17). Estriol was measured by RIA after a dual organic solvent extraction procedure (18). Prolactin, IGF-I, and IGFBP-3 were quantified by direct chemiluminescent immunoassay using the Immulite analyzer (Siemens Medical Solutions Diagnostics, Los Angeles, CA).

Because some of the samples showed signs of hemolysis (34%), measurements were prioritized with the direct assays (that is, prolactin, IGF-I, and IGFBP-3) done in samples with the least hemolysis. After this prioritization, when volume was insufficient to accommodate assays for all of the hormones, samples were randomly assigned based on groupings that optimized the number of assays that could be done. Blinded aliquots of pooled cord sera were included with the study samples and the laboratory technicians were blinded to country of origin. The coefficients of variation for the blinded replicates were 8.5% for androstenedione, 9.4% for testosterone, 10.5% for estradiol, 11.6% for estriol, 8.2% for prolactin, 3.2% for IGF-I, and 5.0% for IGFBP-3.

Information on maternal, gestational, and perinatal characteristics was obtained from the medical record and pediatric chart and from an interview with the mother. Gestational age was defined as the time since the first day of the last menstrual period.

Hormone values that were more than three interquartile ranges above the mean were excluded as outliers. The nonparametric Wilcoxon rank-sum test was applied in univariate comparisons of the maternal, gestational, and perinatal factors between the two study sites (19). Linear regression models with logarithm-transformed hormones as the dependent variable and an indicator variable representing the comparison of Chinese versus Caucasian (and urban versus rural Chinese) were used to generate means for and percent differences in hormone concentrations. Percent difference was calculated as $(\exp^{\beta} - 1) \times 100$, where β pertains to the ethnic/racial comparison, and 95% confidence intervals (95% CI) were calculated as $\exp^{(\beta - 1 \pm 1.96[\mathrm{SE}(\beta)])} \times 100$. Means are geometric (exponentiated from the logarithmic scale). Statistical significance was defined as $P < 0.05$ (two-sided test).

Results

The mothers of Chinese neonates were substantially different from the U.S. mothers on a large number of characteristics. Chinese mothers were significantly

Table 1. Maternal, gestational, and perinatal characteristics among Caucasian women (Boston, MA) and Chinese women (Shanghai, China)

	United States		China		*P*
	n	Mean (range) or %	*n*	Mean (range) or %	
Maternal characteristics					
Age(y)	111	31.2 (19-39)	121	24.9 (20-37)	<0.0001
Weight before pregnancy (kg)	109	60.3 (43-95)	120	51.2 (38-72)	<0.0001
Height (cm)	111	164 (150-183)	121	160 (149-174)	<0.0001
BMI (kg/m^2)	109	22.4 (18-36)	120	20.0 (14-26)	<0.0001
Multiparous	48	43.2	3	2.5	<0.0001
Primiparous	63	56.8	118	97.5	
Gestational characteristics					
Gestational length (wk)	106	40.1 (37-44)	119	39.8 (34-46)	0.31
Weight gained at week 27 (kg)	106	11.5 (3.2-25)	118	9.0 (-3.0 to 22)	<0.0001
Nausea and/or vomiting during pregnancy	87	78.4	77	63.6	0.02
Neither nausea or vomiting during pregnancy	24	21.6	42	34.7	
Perinatal characteristics					
Birth weight (g)	111	3,552 (2,625-4,970)	121	3,463 (1,900-4,800)	0.18
Birth length (cm)	111	50.6 (44-56)	121	49.8 (33-56)	0.10
Head circumference (cm)	109	34.8 (32-38)	119	34.7 (30-51)	0.11
Placenta weight (g)	105	586 (340-1,020)	116	633 (360-1,000)	0.005
Female	58	52.3	51	42.2	0.12
Male	53	47.7	70	57.8	

NOTE: Mean (range) for continuous variables; % for categorical variables.

younger than the U.S. mothers and, on average, were shorter, weighed less before pregnancy, had a lower body mass index (BMI), and gained less weight by the end of the second trimester (Table 1). Chinese mothers were significantly more likely to be primiparous and to have had a C-section (43.3% versus 26.1%, respectively; $P = 0.009$) compared with U.S. mothers and less likely to have completed high school (7.4% versus 97.3%; $P < 0.0001$). In addition, during the pregnancy, Chinese mothers were less likely to experience nausea and/or vomiting (63.6% versus 78.4%; $P = 0.02$), to have drunk coffee (2.5% versus 62.2%; $P < 0.0001$) or tea (12.4% versus 55.9%; $P = 0.0001$), or to have taken antibiotics (3.3% versus 22.5%; $P < 0.0001$). Alcohol consumption during pregnancy was rare among Chinese and U.S. mothers (0.8% versus 3.6%, respectively; $P_{\text{difference}} = 0.20$). There were no statistically significant differences in gender or birth size, including weight, length, and head circumference by race/ethnicity, but placental weight was significantly higher in the Chinese pregnancies. None of the 111 U.S. infants had a birth weight less than 2,500 g or a gestational length less than 37 weeks, and only 3 and 6, respectively, of the 121 Chinese infants were in these categories. There were no statistically significant differences between the Chinese and the U.S. infants, comparing the proportions in the lowest quartiles for anthropometric measurements (based on the distribution in the entire study group). The proportion of Chinese

Table 2. Distributions of cord serum hormone concentrations in U.S. Caucasians (Boston, MA) and Chinese (Shanghai, China)

Hormone	*n*	Mean	5%	25%	Median	75%	95%
Estradiol (nmol/L)							
U.S.	87	34.4	11.7	18.7	28.1	42.3	80.2
Chinese	110	44.9	3.6	15.9	36.6	57.7	125.2
Estriol (nmol/L)							
U.S.	96	541	195	354	460	664	1,135
Chinese	120	829	304	566	814	1,022	1,554
Androstenedione (nmol/L)							
U.S.	88	18.2	7.9	12.7	16.5	21.4	34.3
Chinese	114	43.4	10.9	18.4	28.0	54.9	141
Testosterone (nmol/L)							
U.S.	88	1.1	0.47	0.70	0.92	1.2	2.5
Chinese	115	5.5	0.63	1.5	3.8	7.9	17.3
Prolactin (μg/L)							
U.S.	81	334	145	249	312	436	553
Chinese	47	293	124	200	283	400	487
IGF-I (nmol/L)							
U.S.	51	10.0	4.1	6.0	10.6	13.2	16.2
Chinese	22	11.8	4.4	7.0	11.3	15.4	19.2
IGFBP-3 (nmol/L)							
U.S.	52	32.0	22.3	26.9	31.6	36.3	43.7
Chinese	21	42.7	21.8	27.8	30.5	51.5	89.1

Table 3. Unadjusted means for and percent differences (95% CI) in cord hormone concentrations between U.S. Caucasians (Boston, MA) and Chinese (Shanghai, China)

Hormone	U.S., *n* (mean)	Chinese, *n* (mean)			Chinese vs U.S., % difference (95% CI)	Chinese rural vs urban, % difference (95% CI)
		Total	Urban	Rural		
Estradiol (nmol/L)	87 (28.5)	110 (29.1)	50 (28.7)	60 (29.4)	2.0 (−21.3, 32.3)	2.3 (−32.5, 55.0)
Estriol (nmol/L)	96 (477)	120 (709)	54 (663)	66 (749)	48.6 (23.6, 78.6)	13.0 (−15.3, 50.6)
Androstenedione (nmol/L)	88 (16.7)	114 (32.0)	48 (29.6)	66 (33.8)	91.9 (61.1, 128)	14.3 (−13.5, 50.9)
Testosterone (nmol/L)	88 (0.96)	115 (3.4)	51 (2.6)	64 (4.4)	257 (184, 349)	70.7 (19.4, 144)
Prolactin (μg/L)	81 (311)	47 (265)	14 (244)	33 (274)	−14.9 (−27.0, −0.92)	12.4 (−16.3, 50.8)
IGF-I (nmol/L)	51 (9.1)	22 (10.5)	8 (14.2)	14 (8.8)	14.7 (−9.6, 45.4)	−37.8 (−58.3, −7.2)
IGFBP-3 (nmol/L)	52 (31.2)	21 (37.8)	8 (50.2)	13 (31.8)	21.1 (3.5, 41.8)	−36.6 (−56.5, −7.6)

NOTE: From linear regression models with logarithm-transformed hormones as the dependent variable and Chinese versus U.S. sample (or rural versus urban) as an indicator variable. Percent difference is calculated as $(\exp^{\beta} - 1) \times 100$, where β pertains to ethnic/racial comparison. Geometric means are presented.

infants in the lowest quartile of birth weight was 28.1% versus 21.6% for U.S. infants ($P = 0.29$ for difference in proportions), of head circumference was 15.1% versus 22.9% ($P = 0.18$), and of birth length was 22.3% versus 30.6% ($P = 0.17$), respectively.

The distributions of the hormones are presented in Table 2. Cord serum estriol, androstenedione, testosterone, and IGFBP-3 concentrations were significantly higher in the Chinese than in the U.S. samples (Table 3), whereas prolactin levels were significantly lower. There was no appreciable difference in estradiol concentrations by race/ethnicity, and higher IGF-I levels in the Chinese were not statistically significantly different from U.S. values. Repeating the hormone comparisons between the U.S. and the Chinese samples using Wilcoxon rank-sum tests, the results were similar, except for IGFBP-3, which did not show a statistically significant difference. The patterns of results for percent differences in cord estrogen and androgen concentrations between rural and urban Chinese were in the same direction but of lesser magnitude than those between the U.S. and all Chinese samples combined. However, only the percent difference in testosterone was statistically significant. The sample sizes, especially for prolactin, IGF-I, and IGFBP-3, were small and the 95% CIs were wide.

To determine whether differences in cord hormones between Chinese and U.S. samples were independent of the other hormones, we repeated the analyses adding each of the hormones individually to the models. With adjustment for androgen concentrations, the higher estriol and IGFBP-3 levels observed in the Chinese were attenuated. For example, with androstenedione in the model, the percent difference in estriol decreased from ~49% higher in the Chinese to only 2.2%, and adjusting IGFBP-3 for androstenedione decreased the percent difference from 21% to ~0%. In contrast, the higher androgen levels in the Chinese compared with Caucasians were not affected by adjustment for either of the estrogens or IGFBP-3. Estradiol levels did not differ between Chinese and Caucasians regardless of whether they were unadjusted or adjusted for the other hormones (data not shown).

With adjustment for gestational length, maternal age, pre-pregnancy BMI, and pregnancy weight gain through the second trimester, the androgens and IGFBP-3 remained elevated in the Chinese, but the differences in estriol and prolactin no longer remained statistically significant (Table 4). Percent differences in estriol, prolactin, and the androgens were influenced by adjustment for maternal age, and further adjustment for weight gain also affected percent differences in estriol, prolactin, and IGFBP-3 (data not shown). Estradiol concentrations, which did not differ by race/ethnicity in the unadjusted comparisons, became lower in the Chinese with adjustment (mainly from maternal age), although the racial/ethnic difference did not reach statistical significance. IGF-I levels did not differ by race/ethnicity in either unadjusted or adjusted comparisons. Additional adjustment for maternal height did not change the estimates (data not shown). Adjustment for maternal age was difficult because the Chinese women were quite young and the U.S. women were considerably older with the only appreciable overlap between 24 and 35 years old.

Table 4. Adjusted means for and percent differences (95% CI) in cord serum hormone concentrations between U.S. Caucasians (Boston, MA) and Chinese (Shanghai, China)

Hormone	*n* (Mean)	% difference*
Estradiol (nmol/L)		
U.S.	79 (34.7)	−29.8 (−52.3, 3.5)
Chinese	105 (24.4)	
Estriol (nmol/L)		
U.S.	87 (510)	27.7 (−3.1, 68.3)
Chinese	115 (651)	
Androstenedione (nmol/L)		
U.S.	80 (18.3)	60.5 (23.5, 109)
Chinese	109 (29.4)	
Testosterone (nmol/L)		
U.S.	80 (1.1)	185 (101, 304)
Chinese	110 (3.2)	
Prolactin (μg/L)		
U.S.	73 (301)	−10.1 (−30.6, 16.3)
Chinese	47 (270)	
IGF-I (nmol/L)		
U.S.	45 (9.0)	11.7 (−23.6, 63.2)
Chinese	22 (10.0)	
IGFBP-3 (nmol/L)		
U.S.	46 (29.4)	40.4 (7.8, 82.9)
Chinese	21 (41.3)	

NOTE: From linear regression models with logarithm-transformed hormones as the dependent variable and Chinese versus U.S. samples as an indicator variable. Percent difference is calculated as $(\exp^{\beta} - 1) \times 100$, where β pertains to ethnic/racial comparison. Geometric means are presented.

*Model includes gestational length, maternal age, pre-pregnancy BMI, and weight gain as independent variables.

Repeating the analyses in this age range (24-35 years) and controlling for age as a continuous variable, the results were similar to those observed in the overall group (data not shown). Additional adjustment for placental weight did not change the results, although the higher estriol concentrations among Chinese neonates increased from ~28% to 38% (95% CI, 2.5%, 85%). Furthermore, adding offspring gender to the adjusted models did not change the percent differences in the overall analysis (estradiol, −31.8% with offspring gender versus −29.8% without offspring gender; estriol, 24.4% versus 27.7%; androstenedione, 53.8% versus 60.5%; testosterone, 166% versus 185%; prolactin, −10.7% versus −10.1%; IGF-I, 14.3% versus 11.7%; IGFBP-3, 41.7% versus 40.4%).

Nearly all of the Chinese women were primiparous; therefore, we restricted analyses to neonates born to women with no previous live or still births. The pattern of unadjusted and adjusted results was generally similar to the overall differences with significantly elevated estriol, androstenedione, testosterone, and IGFBP-3 concentrations in the Chinese. For example, the unadjusted and adjusted (for maternal age, pre-pregnancy weight, and weight gain) percent differences in estriol in all women were 48.6% and 27.7% versus 47.5% and 22.0% in primiparous women. The corresponding values were 91.9% and 60.5% versus 78.5% and 56.4% for androstenedione and 257% and 185% versus 225% and 162% for testosterone. For estradiol, the unadjusted values were similar, whereas the adjusted differences became somewhat greater and achieved statistical significance [unadjusted percent difference, −10.4% (95% CI, −35.4%, 24.2%); adjusted percent difference, −45.2% (95% CI, −65.2%, −13.6%)]. As the results for hormones by race/ethnicity were in the same direction and remained statistically significant regardless of parity, we included multiparous women in the overall analysis to increase statistical power. When stratified by offspring gender, the magnitude of the percent differences was generally greater in males than females, but these comparisons were limited by the reduction in sample sizes. For example, the percent differences were 98.3% (95% CI, 53.7%, 156%) for androstenedione and 276% (95% CI, 178%, 408%) for testosterone in the male infants and 74.5% (95% CI, 38.8%, 120%) and 213% (95% CI, 123%, 339%) among the females, respectively. None of the interactions of the associations of hormones and race/ethnicity by offspring gender were statistically significant, ranging from 0.20 for IGF-I to 0.91 for prolactin.

Discussion

The data presented here are the first to directly address differences in fetal hormone concentrations between pregnancies occurring in geographic regions characterized by relatively low (China) and high (United States) breast cancer incidence rates. The elevated cord androgen concentrations we observed, however, are consistent with previous studies of pregnancies occurring in North America showing higher dehydroepiandrosterone sulfate (DHEAS) concentrations in Chinese Canadians compared with Caucasians (20), as is the pattern of higher cord estriol concentrations that we observed in the Chinese (20). In contrast, estradiol concentrations were not elevated in the Chinese neonates in our study. In a previous study, cord estradiol concentrations were higher in Chinese American than in Caucasian pregnancies (21) with or without adjustment for several pregnancy factors, although the study was small and consisted of women recruited from clients of a cord blood registry, suggesting relatively high acculturation. We found no difference in IGF-I by race/ethnicity in our data, whereas IGFBP-3 levels were elevated in the Chinese. The study cited above (21) reported no differences in unadjusted IGF profiles, including IGF-I and IGFBP-3 levels, comparing Chinese American with U.S. pregnancies.

Our findings for differences in cord estrogen and prolactin levels between Chinese and U.S. neonates were not entirely similar to hormone differences observed in the maternal data for these same pregnancies. Whereas androgen and IGF concentrations were not assessed in the study of serum hormones in the mothers of these infants (13), maternal levels of estradiol and prolactin as well as estriol were elevated in the Chinese. The vast majority (over 90%) of estradiol and estriol enter the maternal compartment from the placenta (22). Most studies have measured hormones and other biomarkers in the maternal circulation due to the difficulty in directly sampling the *in utero* environment and based on the assumption that maternal hormones and other endocrine factors reflect those in the fetal circulation because of the highly integrated maternal-placental-fetal unit. However, the degree of correlation between, for example, estrogen and androgen concentrations in the maternal and fetal circulations is modest (23) and may explain the differences in the direction of results for the maternal and cord samples in the present data.

We are not aware of studies that have measured fetal androgen concentrations at multiple points throughout the pregnancy. Whether hormone levels in the fetal circulation reflect levels in breast tissue is not known and is an issue in any study that uses such proxies for fetal exposure. In epidemiologic studies, umbilical cord sampling is only feasible after delivery; thus, the differences we observed between groups may not represent real differences earlier in pregnancy. In particular, androgen concentrations may be higher after vaginal deliveries than after C-sections. However, the proportion of C-sections was actually greater in the Chinese than U.S. women (41.3% versus 26.1%) and thus would not explain the higher androgen concentrations we observed in the Chinese cord samples.

We observed elevated androgen and estriol concentrations but either no difference (unadjusted) or reductions (adjusted) in estradiol in the Chinese. In uncomplicated pregnancies, nearly comparable amounts of DHEAS from the maternal and fetal adrenal glands are enzymatically converted in the placenta to androstenedione and testosterone, which are then aromatized to estrone and estradiol, respectively (22). The conversion rate of androgens to estradiol is a function of placental size and capacity as well as of aromatase enzyme levels. Estriol is synthesized by the placenta from 16α-hydroxy-DHEAS, which is formed in the fetal liver from DHEAS. Over 90% of urinary estriol are ultimately derived from the fetal adrenal gland (22). As pregnancy estriol is derived from androgen substrate, these hormones are necessarily correlated (r = 0.39 for androstenedione and estriol and r = 0.30 for testosterone and estriol in the present study). Thus, the higher estriol concentrations in the Chinese that

we observed could merely be due to higher androgen concentrations, which were also noted in the Chinese. When androstenedione was added to the regression model for estriol, the higher estriol concentrations in the Chinese infants were markedly attenuated, whereas androstenedione remained significantly higher in the Chinese. This suggests that the elevated estriol levels in the Chinese are probably due to their greater amounts of estrogen precursor (that is, fetal androgens). The attenuation in the group differences for estriol with androstenedione in the model could also result if androstenedione was measured with less laboratory error than estriol. However, the coefficients of variation were fairly similar for the two hormones.

The similar or reduced estradiol levels in the presence of elevated testosterone concentrations are more difficult to explain. This implies less aromatization in the Chinese, but it does not appear to be explained by placental size as average placental weight was higher in the Chinese pregnancies. Incomplete aromatization in the placental compartment could also explain the higher testosterone concentrations in the Chinese. Alternatively, or in addition to this explanation, the higher androgens in the Chinese may be due to greater concentrations in the fetal compartment. In this regard, the hormone results by offspring gender may add to the understanding of the biology of the higher androgen levels in the Chinese. Given the values are higher in the Chinese regardless of gender implies that placental differences (that is, in aromatization) between Chinese and U.S. pregnancies may be responsible.

Our unadjusted data would not appear to support the hypothesis that exposure to lower estrogen levels *in utero* are responsible in part for the lower breast cancer risk in Chinese women. Other prenatal factors related to breast cancer risk have been proposed to be mediated through differences in pregnancy estrogen exposure, including high birth weight (24) and dizygotic twinning (25-30) as well as the reduced risk observed for women born of preeclamptic pregnancies (31). Birth weight has been positively associated with maternal estrogens in several studies (32-34). However, data validating the associations of preeclampsia (34, 35) and birth weight (21, 36-38) with estrogen levels, particularly in the cord, are conflicting, and data for dizygotic twinning are lacking (39).

We hypothesize that a difference in fetal androgen exposure at a critical period during pregnancy may explain the lower breast cancer rates in Asians compared with Caucasians. Androgen concentrations in the cord circulation of the Chinese neonates were two to three times greater than in the U.S. neonates. Elevated fetal androgen concentrations have been proposed as mediating the associations of prenatal exposures with breast cancer risk (40) possibly through reduction of the initial breast stem cell population. Suppression of embryonic mammary gland development by androgens in males supports this hypothesis (41). In the mouse model, destruction of mammary gland anlagen (the initial clustering of cells destined to become breast tissue) by testosterone occurs in early pregnancy, and this androgen sensitivity is expressed in male as well as female mammary glands (42). Female fetuses are protected from sterilization through rapid placental metabolism of androgens to estrogens (43). Given that the prohibitive effects of androgens on breast anlagen formation occur in female as well as male mammary glands and the empirical evidence that women undergo breast development whereas men generally do not, we believe rapid androgen metabolism plays a protective role in allowing female breast development as well as in preventing virilization. Androgen variability below the level of sterilization, however, may have significant biological consequences. Thus, it is possible that limited androgen transfer back to the fetus in cases in which androgen levels are high could influence anlagen formation, as it does in males. Whether the magnitude of the difference in androgen concentrations between Chinese and U.S. infants that we observed is sufficient to protect the breast is unknown. Elevated androgen concentrations are observed for other prenatal exposures that are associated with a reduced breast cancer risk, such as preeclampsia (34, 35). These observations would be consistent with a protective effect of fetal androgens but the data, particularly for cord concentrations, are sparse.

Associations of the hormones studied with other maternal and pregnancy factors could explain the differences observed by race/ethnicity. In particular, the Chinese mothers tended to be younger and physically smaller with lower pre-pregnancy weight, height, and BMI and less pregnancy weight gain. The differences in androgens and IGFBP-3 remained with adjustment for these factors, and those for estriol and prolactin were attenuated. In contrast, the estradiol concentrations became lower than in Caucasians following adjustment. The similar estradiol concentrations in the two racial/ethnic groups in unadjusted comparisons were largely due to younger maternal age in Chinese women.

We presented both unadjusted means and those adjusted for maternal and pregnancy factors shown previously to be associated with cord hormone levels and which were related to race/ethnicity in these data. We believe the unadjusted results address whether cord hormones explain the international difference in breast cancer rates in offspring, rate differences that are unadjusted for maternal characteristics. For example, if estradiol concentrations are similar in the Chinese and U.S. mothers because the Chinese mothers tend to be younger and smaller, then (assuming our population samples are representative and Chinese women are indeed generally younger and smaller) estradiol seems unlikely to explain the international rate differences. The unadjusted results also could be more relevant to the actual pregnancy exposure of the fetus. In subsequent analyses, we adjusted for maternal factors, including age, pre-pregnancy BMI and weight gain, and length of the gestation.

The adjusted results more appropriately address whether the hormone differences we observed between Chinese and U.S. infants are due to differences in maternal age and size. Some but not all of the higher androgen levels in the Chinese appear to be due to their younger maternal age. For estriol and prolactin, the higher levels in the Chinese appear to be due to their younger maternal age and to less pregnancy weight gain. The lack of difference in the crude data for estradiol seems to be driven by the younger maternal age of the Chinese. If either androgen or estrogen levels are causally responsible for the international differences in breast cancer risk, it would seem that their correlates (that is, maternal age) would be risk factors for breast

cancer. However, until now, this has not been consistently observed (44). Regardless of which results are used, the unadjusted or adjusted, and focusing only on fetal exposure, higher androgens and IGFBP-3 are consistent with protective effects on breast cancer risk, whereas prolactin is consistent with adverse effects.

The samples of women from China and the United States were not population based, and as such, the data may not be representative of pregnancies occurring in each country. The Chinese women, however, were mainly from Shanghai and likely had a more western lifestyle than would be typical in China. Thus, any observed hormone differences that were due to environmental factors may be underestimated. In fact, the hormone profile for the urban Chinese, although closer to the rural Chinese, was between that of the Caucasians and that of rural Chinese. Some of the cord sera showed signs of hemolysis and could not be used for the direct assays (that is, prolactin, IGF-I, and IGFBP-3), resulting in smaller sample sizes for these analytes and less power to detect differences by race/ethnicity.

In conclusion, we found higher concentrations of estriol, androstenedione, testosterone, and IGFBP-3 and lower prolactin in cord serum from Chinese compared with U.S. neonates, whereas levels of estradiol and IGF-I were not different. These data are consistent with the hypothesis that elevated prenatal androgen exposure may be protective against subsequent breast carcinogenesis. Whereas the focus of this paper is on breast cancer, these results may be relevant for other endocrine cancers that have been suggested as being associated with fetal hormone exposure, including testicular and prostate cancer. Given that cultural and lifestyle practices are changing in parts of Asia, studying populations, for example, that have moved from rural to urban areas to address whether changes in lifestyle factors affect pregnancy-maternal and umbilical cord-hormone concentrations could be useful.

References

1. CANCER*Mondial*; 2002. Available from: http://www-dep.iarc.fr/.
2. Ziegler RG, Hoover RN, Pike MC, et al. Migration patterns and breast cancer risk in Asian-American Women. J Natl Cancer Inst 1993;85:1819–27.
3. Wu AH, Ziegler RG, Pike MC, et al. Menstrual and reproductive factors and risk of breast cancer in Asian-Americans. Br J Cancer 1996;73:680–6.
4. Ursin G, Wu AH, Hoover RN, et al. Breast cancer and oral contraceptive use in Asian-American women. Am J Epidemiol 1999;150:561–7.
5. Buell P. Changing incidence of breast cancer in Japanese-American women. J Natl Cancer Inst 1973;51:1479–83.
6. Miller AB, Bulbrook RD. UICC Multidisciplinary Project on Breast Cancer: the epidemiology, aetiology and prevention of breast cancer. Int J Cancer 1986;37:173–7.
7. Willett WC, MacMahon B. Diet and cancer—an overview. N Engl J Med 1984;310:633–8, 697–703.
8. National Cancer Institute. Summary report: early reproductive events and breast cancer workshop; 2003. Available from: http://cancer.gov/cancerinfo/ereworkshop-report.
9. Trichopoulos D. Hypothesis: does breast cancer originate *in utero*? Lancet 1990;335:939–40.
10. Key TJA, Chen J, Wang DY, Pike MC, Boreham J. Sex hormones in women in rural China and in Britain. Br J Cancer 1990;62:631–6.
11. Bernstein L, Yuan J-M, Ross RK, et al. Serum hormone levels in premenopausal Chinese women in Shanghai and white women in Los Angeles: results from two breast cancer case-control studies. Cancer Causes Control 1990;1:51–8.
12. Falk RT, Fears TR, Hoover RN, et al. Does place of birth influence endogenous hormone levels in Asian-American women. Br J Cancer 2002;87:54–60.
13. Lipworth L, Hsieh C-C, Wide L, et al. Maternal pregnancy hormone levels in an area with a high incidence (Boston, USA) and in an area with a low incidence (Shanghai, China) of breast cancer. Br J Cancer 1999;79:7–12.
14. Lambe M, Trichopoulos D, Hsieh CC, Wuu J, Adami HO, Wide L. Ethnic differences in breast cancer risk: a possible role for pregnancy levels of alpha-fetoprotein? Epidemiology 2003;14:85–9.
15. Goebelsmann U, Bernstein GS, Gale JA, et al. Serum gonadotropin, testosterone, estradiol and estrone levels prior to and following bilateral vasectomy. In: Lepow IH, Crozier R, eds. Vasectomy: immunologic and pathophysiologic effects in animals and man. New York: Academic Press; 1979. p. 165.
16. Goebelsmann U, Arce JJ, Thorneycroft IH, Mishell DR, Jr. Serum testosterone concentrations in women throughout the menstrual cycle and following hCG administration. Am J Obstet Gynecol 1974;119:445–52.
17. Goebelsmann U, Horton R, Mestman JH, et al. Male pseudohermaphroditism due to testicular 17-hydroxysteroid dehydrogenase deficiency. J Clin Endocrinol Metab 1973;36:867.
18. Katagiri H, Stanczyk FZ, Goebelsmann U. Estriol in pregnancy. III. Development, comparison and use of specific antisera for rapid radioimmunoassay of unconjugated estriol in pregnancy plasma. Steroids 1974;24:225–38.
19. Armitage P, Berry G. Statistical methods in medical research. Boston: Blackwell Boston; 1987.
20. Yeung LPK, Wong ACK, Wang X, Birmingham CL, Lewicka S, Chanoine J-P. Different relationship between anthropometric markers and umbilical cord plasma leptin in Asian and Caucasian neonates. Pediatric Res 2003;53:1019–24.
21. Shibata A, Harris DT, Billings PR. Concentrations of estrogens and IGFs in umbilical cord blood plasma: a comparison among Caucasian, Hispanic, and Asian-American females. J Clin Endocrinol Metab 2002;87:810–5.
22. Carr BR. The maternal-fetal-placental unit. In: Becker KL, editor. Principles and practice of endocrinology and metabolism. 2nd ed. Philadelphia: J.B. Lippincott; 1995.
23. Troisi R, Potischman N, Roberts JM, et al. Correlation of serum hormone concentrations in maternal and umbilical cord samples. Cancer Epidemiol Biomarkers Prev 2003;12:452–6.
24. Lagiou P, Samoli E, Lagiou A, Hsieh CC, Adami HO, Trichopoulos D. Maternal height, pregnancy estriol and birth weight in reference to breast cancer risk in Boston and Shanghai. Int J Cancer 2005;117:494–8.
25. Ekbom A, Hsieh C-C, Lipworth L, Adami H-O, Trichopoulos D. Intrauterine environment and breast cancer risk in women: a population-based study. J Natl Cancer Inst 1997;88:71–6.
26. Hsieh C-C, Lan S-J, Ekbom A, Petridou E, Adami H-O, Trichopoulos D. Twin membership and breast cancer risk. Am J Epidemiol 1992;136:1321–6.
27. Weiss HA, Potischman NA, Brinton LA, et al. Prenatal and perinatal risk factors for breast cancer in young women. Epidemiology 1997;8:181–7.
28. Braun MM, Ahlbom A, Floderus B, Brinton L, Hoover RN. Effect of twinship on incidence of cancer of the testis, breast and other sites (Sweden). Cancer Causes Control 1995;6:519–24.
29. Cerhan JR, Kushi LH, Olson JE, et al. Twinship and risk of postmenopausal breast cancer. J Natl Cancer Inst 2000;92:261–5.
30. Verkasalo PK, Kaprio J, Pukkala E, Koskenvuo M. Breast cancer risk in monozygotic and dizygotic female twins: a 20-year population-based cohort study in Finland from 1976 to 1995. Cancer Epidemiol Biomarkers Prev 1999;8:271–4.
31. Innes KE, Byers TE. Preeclampsia and breast cancer risk. Epidemiology 1999;10:722–32.
32. Mucci LA, Lagiou P, Tamimi RM, Hsieh CC, Adami H-O, Trichopoulos D. Pregnancy estriol, estradiol, progesterone and prolactin in relation to birth weight and other birth size variables (United States). Cancer Causes Control 2003;14:311–8.
33. Kaijser M, Granath F, Jacobsen G, Cnattingius S, Ekbom A. Maternal pregnancy estriol levels in relation to anamnestic and fetal anthropometric data. Epidemiology 2000;11:315–9.
34. Troisi R, Potischman N, Neuser Johnson C, et al. Estrogen and androgen concentrations are not lower in the umbilical cord serum of preeclamptic pregnancies. Cancer Epidemiol Biomarkers Prev 2003;12:1268–70.
35. Atamer Y, Erden AC, Demir B, Koçyigit Y, Atamer A. The relationship between plasma levels of leptin and androgen in healthy and preeclamptic pregnant women. Acta Obstet Gynecol Scand 2004;83:425–30.

36. Troisi R, Potischman N, Roberts JM, Siiteri P, Hoover RN. Associations of maternal and umbilical cord hormone concentrations with maternal, gestational and neonatal factors. Cancer Causes Control 2003;14:347–55.
37. Nagata C, Iwasa S, Shiraki M, Shimizu H. Estrogen and α-fetoprotein levels in maternal and umbilical cord blood samples in relation to birth weight. Cancer Epidemiol Biomarkers Prev 2006; 15:1469–72.
38. Simmons D, France JT, Keelan JA, Song L, Knox BS. Sex differences in umbilical cord serum levels of inhibin, testosterone, oestradiol, dehydroepiandrosterone sulphate, and sex hormone-binding globulin in human term neonates. Biol Neonate 1994;65:287–94.
39. Thomas HV, Murphy MF, Key TJ, Fentiman IS, Allen DS, Kinlen LJ. Pregnancy and menstrual hormone levels in mothers of twins compared to mothers of singletons. Ann Hum Biol 1998;25:69–75.
40. Troisi R, Potischman N, Roberts JM, et al. Maternal serum hormone concentrations in preeclamptic and uncomplicated pregnancies. Int J Epidemiol 2003;32:455–60.
41. Drews U, Drews U. Regression of mouse mammary gland anlagen in recombinants of *Tfm* and wild-type tissues: testosterone acts via the mesenchyme. Cell 1977;10:401–4.
42. Wald N, Cuckle H, Wu TS, George L. Maternal serum unconjugated oestriol and human chorionic gonadotrophin levels in twin pregnancies: implications for screening for Down's syndrome. Br J Obstet Gynaecol 1991;98:905–8.
43. Yen SSC. Endocrine-metabolic adaptations in pregnancy. In: Yen SSC, Jaffe RB, editors. Reproductive endocrinology. Philadelphia: W.B. Saunders; 1991. p. 936–81.
44. Potischman N, Troisi R. *In-utero* and early life exposures in relation to risk of breast cancer. Cancer Causes Control 1999;10:561–73.